



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>C12N 15/12, C07K 14/705, C12Q 1/68, G01N 33/68, A61K 38/16, 35/00</b>		<b>A2</b>	(11) International Publication Number: <b>WO 99/46380</b>
			(43) International Publication Date: 16 September 1999 (16.09.99)
(21) International Application Number: PCT/US99/05073 (22) International Filing Date: 9 March 1999 (09.03.99) (30) Priority Data: 09/039,064                      13 March 1998 (13.03.98)                      US (63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application US    09/039,064 (CIP) Filed on                                      13 March 1998 (13.03.98) (71) Applicant (for all designated States except US): INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). BAND- MAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 9054 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #12, Mountain View, CA 94040 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sun- nyvale, CA 94087 (US). CORLEY, Neil, C. [US/US]; 1240		Dale Avenue #30, Mountain View, CA 94040 (US). GUE- GLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). KASER, Matthew, R. [GB/US]; 4793 Ewing Road, Castro Valley, CA 94546-1017 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). SHAH, Purvi [IN/US]; 1608 Queen Charlotte Drive #5, Sunnyvale, CA 94087 (US). (74) Agents: BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US). (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published Without international search report and to be republished upon receipt of that report.	
(54) Title: HUMAN MEMBRANE SPANNING PROTEINS			
(57) Abstract			
The invention provides a human membrane spanning proteins (MSPs) and polynucleotides which identify and encode MSPs. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for treating or preventing disorders associated with expression of MSPs.			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## HUMAN MEMBRANE SPANNING PROTEINS

### TECHNICAL FIELD

5           This invention relates to nucleic acid and amino acid sequences of human membrane spanning proteins and to the use of these sequences in the diagnosis, treatment, and prevention of neoplastic, immunological, and reproductive disorders.

### BACKGROUND OF THE INVENTION

10           Eukaryotic organisms are distinct from prokaryotes in possessing many intracellular organelle and vesicle structures. Many of the metabolic reactions which distinguish eukaryotic biochemistry from prokaryotic biochemistry take place within these structures. In particular, many cellular functions require very strict reagent conditions, and the organelles and vesicles enable compartmentalization and isolation of reactions which  
15 might otherwise cripple cytosolic metabolic processes. The organelles are bounded by single or double membranes each made up of lipid bilayer sheets and include mitochondria; smooth and rough endoplasmic reticuli; sarcoplasmic reticulum; and the Golgi body. The lipid bilayer sheets are composed of phosphoglycerides, fatty acids, cholesterol, glycolipids, proteoglycans, and proteins. The vesicles are bounded by single  
20 membranes and include phagosomes; lysosomes; endosomes; peroxisomes; and secretory vesicles.

          Biological membranes are highly selective permeability barriers that contain ion pumps, gates, and specific receptors for external stimuli which transmit biochemical signals across the membranes. These membranes also contain proteins which interact with  
25 these pumps, gates, and receptors to amplify and regulate transmission of these signals.

#### **Plasma Membrane Proteins**

          Plasma membrane proteins (MPs) are divided into two groups based upon the mechanism of protein removal from the membrane. Extrinsic or peripheral membrane proteins can be removed using extremes of ionic strength or pH, urea, or other disruptors  
30 of protein interactions. Intrinsic or integral membrane proteins are released only when the lipid bilayer of the membrane is dissolved by detergent.

The majority of known integral membrane proteins are transmembrane proteins which are characterized by an extracellular, a transmembrane, and an intracellular domain. Transmembrane proteins are typically embedded in the cell membrane by one or more regions comprising 15 to 25 hydrophobic amino acids which are predicted to adopt an  $\alpha$ -helical conformation. Transmembrane proteins are classified as bitopic (Types I and II) and polytopic (Types III and IV). (Singer, S.J. (1990) *Annu. Rev. Cell Biol.* 6:247-96.) Bitopic proteins span the membrane once while polytopic proteins contain multiple membrane-spanning segments. Type III integral membrane proteins have multiple transmembrane stretches of hydrophobic residues. Transmembrane proteins carry out a variety of important cellular functions such as acting as cell-surface receptor proteins involved in signal transduction, e.g., growth factor and differentiation factor receptors and receptor-interacting proteins; transport of ions or metabolites, e.g., gap junction channels (connexins), and ion channels; cell anchoring proteins, extracellular matrix (ECM)-binding proteins, lectins, integrins, and fibronectins; or cell recognition molecules, e.g., cluster of differentiation (CD) antigens, glycoproteins and mucins.

Many MPs contain amino acid sequence motifs that serve to localize proteins to specific subcellular sites. Examples of these motifs include, e.g., PDZ domains, KDEL, RGD, NGR, GSL, von Willebrand factor A (vWFA) domains, and EGF-like domains. Furthermore, MPs may also contain amino acid sequence motifs that serve to interact with extracellular or intracellular molecules, such as carbohydrate recognition domains (CRD). RGD, NGR, GSL motif-containing peptides have been used as drug delivery agents in targeted cancer treatment of tumor vasculature. (Arap, W. et al. (1998) *Science*, 279:377-380.)

Chemical modification of amino acid residue side chains alters the manner in which a protein interacts with other molecules, for example, phospholipid membranes. Examples of such chemical modifications to amino acid residue side chains are covalent bond formation with glycosaminoglycans, oligosaccharides, phospholipids, acetyl and palmitoyl moieties, ADP-ribose, phosphate, and sulphate groups.

Premessenger RNA encoding membrane proteins may have alternative splice sites which give rise to proteins encoded by the same gene but with different messenger RNA and amino acid sequences. Splice variant membrane proteins may interact with other ligand and protein isoforms.

### Tetraspan family proteins

The transmembrane 4 superfamily (TM4SF) or tetraspan family are a multigene family encoding type III integral membrane proteins. (Wright, M.D. and Tomlinson, M.G. (1994) Immunol. Today 15:588.) The TM4SF family comprises a superfamily of

5 membrane proteins which traverse the cell membrane four times. Members of the TM4SF include a number of platelet and endothelial cell membrane proteins, the platelet and melanoma-associated antigens, leukocyte surface glycoproteins, the colonal carcinoma antigens, the tumor-associated antigens, and surface proteins of the schistosome parasites. (Jankowski, S.A. (1994) Oncogene 9:1205-1211.) The members of the TM4SF share

10 about 25-30% amino acid sequence identity with one another.

A number of TM4SF members have been implicated in signal transduction, control of cell adhesion, regulation of cell growth and proliferation, including development and oncogenesis, and motility, including the ability to suppress metastatic potential. Expression of a number of TM4SF proteins is associated with a variety of tumors and the

15 level of expression may be altered when cells are growing or activated.

### Proton pumps

Proton ATPases are a large class of membrane-proteins that use the energy of ATP hydrolysis to generate an electrochemical proton gradient across a membrane. The resultant gradient may be used to transport other ions across the membrane ( $\text{Na}^+$ ,  $\text{K}^+$ , or  $\text{Cl}^-$

20 ) or to maintain an acidic environment important to the function of many cellular vesicles. (Mellman, I. et al. (1986) Ann. Rev. Biochem. 55:663-700.) Proton ATPases are further subdivided into the mitochondrial F-ATPases, the plasma membrane ATPases, and the vacuolar ATPases. The vacuolar ATPases establish and maintain an acidic pH within various vesicles involved in the processes of endocytosis and exocytosis.

25 Proton-coupled, 12 membrane-spanning domain transporters such as PEPT 1 and PEPT 2 are responsible for gastrointestinal absorption and for renal reabsorbtion of peptides using an electrochemical  $\text{H}^+$  gradient as the driving force. A heterodimeric peptide transporter, consisting of TAP 1 and TAP 2, is associated with antigen processing. Peptide antigens are transported across the membrane of the endoplasmic reticulum so

30 they can be presented to the major histocompatibility complex class I molecules. Each TAP protein consists of multiple hydrophobic membrane spanning segments and a highly conserved ATP-binding cassette. (Boll, M. et al. (1996) Proc. Natl. Acad. Sci.

93:284-289.) Pathogenic microorganisms, such as herpes simplex virus, may encode inhibitors of TAP-mediated peptide transport in order to evade immune surveillance.

(Marusina, K. and Manaco, J.J. (1996) Curr. Opin. Hematol. 3:19-26.)

#### Scavenger receptors

5           Macrophage scavenger receptors with broad ligand specificity have been suggested to take part in the binding of low density lipoproteins (LDL) and foreign antigens.

Scavenger receptors types I and II are trimeric membrane proteins with a small N-terminal intracellular domain, a transmembrane domain, a large extracellular domain, and a C-terminal cysteine-rich domain. The extracellular domain contains a short spacer domain,  
10 an  $\alpha$ -helical coiled-coil domain, and a triple helical collagenous domain. These receptors have been shown to bind a spectrum of ligands, including chemically modified lipoproteins and albumin, polyribonucleotides, polysaccharides, phospholipids, and asbestos. (Matsumoto, A. et al. (1990) Proc. Natl. Acad. Sci. 87:9133-9137; and Elomaa, O. et al. (1995) Cell 80:603-609.)

15           The scavenger receptors are thought to play a key role in atherogenesis by mediating uptake of modified LDL in arterial walls, and in host defense by binding bacterial endotoxins, bacteria, and protozoa.

#### Ion channels

Ion channels are found in the plasma membranes of virtually every cell in the body.  
20 For example, chloride channels mediate a variety of cellular functions including regulation of membrane potentials and absorption and secretion of ions across epithelial membranes. When present in intracellular membranes of the Golgi apparatus and endocytic vesicles, chloride channels also regulate organelle pH. (see, e.g., Greger, R. (1988) Annu. Rev. Physiol. 50:111-122.) Electrophysiological and pharmacological measurements including  
25 ion conductance, current-voltage relationships, and sensitivity to modulators suggest that different chloride channels exist in muscles, neurons, fibroblasts, epithelial cells, and lymphocytes.

Many of the channels have sites for phosphorylation by one or more protein kinases including protein kinase A, protein kinase C, tyrosine kinase, and casein kinase II,  
30 all of which regulate ion channel activity in cells. Inappropriate phosphorylation of proteins in cells has been linked to changes in cell cycle events and differentiation status. Changes in the cell cycle have been linked to induction of apoptosis or cancer. Changes in

differentiation status of the cell have been linked to, for example, immune diseases and disorders, diseases and disorders of skeletal muscle, and diseases and disorders of the reproductive system.

#### G-Protein Coupled Receptors

5 G-protein coupled receptors (GPCR) are a superfamily of integral membrane proteins which transduce extracellular signals. GPCRs include receptors for biogenic amines; for lipid mediators of inflammation, peptide hormones, and sensory signal mediators.

The structure of these highly-conserved receptors consists of seven hydrophobic  
10 transmembrane (serpentine) regions, cysteine disulfide bridges between the second and third extracellular loops, an extracellular N-terminus, and a cytoplasmic C-terminus. Three extracellular loops alternate with three intracellular loops to link the seven transmembrane regions. The most conserved parts of these proteins are the transmembrane regions and the first two cytoplasmic loops. A conserved,  
15 acidic-Arg-aromatic residue triplet present in the second cytoplasmic loop may interact with the G proteins. The consensus pattern of the G-protein coupled receptors signature (PS00237; SWISSPROT) is characteristic of most proteins belonging to this superfamily. (Watson, S. and S. Arkinstall (1994) The G-protein Linked Receptor Facts Book, Academic Press, San Diego, CA, pp 2-6.)

20 Mutations and changes in transcriptional activation of G protein-encoding genes have been associated with, for example, schizophrenia, Parkinson's disease, Alzheimer's disease, drug addiction, and feeding disorders.

#### ABC Transporters

The ATP-binding cassette (ABC) transporters, also called the "traffic ATPases",  
25 comprise a superfamily of membrane proteins that mediate transport and channel functions in prokaryotes and eukaryotes. (Higgins, C.F. (1992) *Annu. Rev. Cell Biol.* 8:67-113.) ABC proteins share a similar overall structure and significant sequence homology. All ABC proteins contain a conserved domain of approximately two hundred amino acid residues which includes one or more nucleotide binding domains. Mutations in ABC  
30 transporter genes are associated with, for example, hyperbilirubinemia II/Dubin-Johnson syndrome, recessive Stargardt's disease, X-linked adrenoleukodystrophy, multidrug resistance, celiac disease, and cystic fibrosis.

### Tumor Antigens

Tumor antigens are surface molecules that are differentially expressed in tumor cells relative to non-tumor tissues. Tumor antigens distinguish tumor cells immunologically from normal cells and provide diagnostic and therapeutic targets for human cancers. (Takagi, S. et al. (1995) Int. J. Cancer 61: 706-715; Liu, E. et al. (1992) Oncogene 7: 1027-1032.)

### **Mitochondrial Membrane Proteins**

The mitochondrial electron transport (or respiratory) chain is a series of three enzyme complexes in the mitochondrial membrane that is responsible for the transport of electrons from NADH to oxygen and the coupling of this oxidation to the synthesis of ATP (oxidative phosphorylation). ATP then provides the primary source of energy for driving the many energy-requiring reactions of a cell.

Most of the protein components of the mitochondrial respiratory chain are the products of nuclear encoded genes that are imported into the mitochondria, and the remainder are products of mitochondrial genes. Defects and altered expression of enzymes in the respiratory chain are associated with a variety of disease conditions in man, including, for example, neurodegenerative diseases, myopathies, and cancer.

### **Endoplasmic Reticulum Membrane Proteins**

The normal functioning of the eukaryotic cell requires that all newly synthesized proteins be correctly folded, modified, and delivered to specific intra- and extracellular sites. Newly synthesized membrane and secretory proteins enter a cellular sorting and distribution network during or immediately after synthesis (cotranslationally or posttranslationally) and are routed to specific locations inside and outside of the cell. The initial compartment in this process is the endoplasmic reticulum (ER) where proteins undergo modifications such as glycosylation, disulfide bond formation, and assembly into oligomers. The proteins are then transported through an additional series of membrane-bound compartments which include the various cisternae of the Golgi complex, where further carbohydrate modifications occur. Transport between compartments occurs by means of vesicles that bud and fuse in a specific manner. Once within the secretory pathway, proteins do not have to cross a membrane to reach the cell surface.

The majority of proteins processed through the ER are transported out of the organelle however some are retained. The signal for retention in the ER in mammalian



cells consists of the tetrapeptide sequence, KDEL, located at the carboxy terminus of proteins. (Munro, S. (1986) Cell 46:291-300.) Proteins containing this sequence leave the ER but are quickly retrieved from the early Golgi compartment and returned to the ER, while proteins without this signal continue through the distribution pathway.

- 5           Disruptions in the cellular secretory pathway have been implicated in several human diseases. In familial hypercholesterolemia the low density lipoprotein receptors remain in the ER, rather than moving to the cell surface. (Pathak, R.K. (1988) J. Cell Biol. 106:1831-1841.)           Presenilins are localized to the ER and may regulate cellular calcium homeostasis in early-onset Alzheimer's disease and in Down syndrome.
- 10       Changes in ER-derived calcium homeostasis have also been associated with, for example, cardiomyopathy, cardiac hypertrophy, myotonic dystrophy, Brody disease, Smith-McCort dysplasia, and diabetes mellitus.

#### **Intercellular Communication Membrane Proteins**

- Intercellular communication is essential for the development and survival of
- 15   multicellular organisms. Cells communicate with one another through the secretion and uptake of protein signaling molecules. The uptake of proteins into the cell is achieved by endocytosis, in which the interaction of signaling molecules with the plasma membrane surface, often via binding to specific receptors, results in the formation of plasma membrane-derived vesicles that enclose and transport the molecules into the cytosol. The
- 20   secretion of proteins from the cell is achieved by exocytosis, in which molecules inside of the cell are packaged into membrane-bound transport vesicles derived from the *trans* Golgi network. These vesicles fuse with the plasma membrane and release their contents into the surrounding extracellular space. Endocytosis and exocytosis result in the removal and addition of plasma membrane components, and the recycling of these components is
- 25   essential to maintain the integrity, identity, and functionality of both the plasma membrane and internal membrane-bound compartments.

- Endocytosis involves the internalizing of nutrients, solutes or small particles (pinocytosis); or large particles such as internalized receptors, viruses, bacteria, or bacterial toxins (phagocytosis). In exocytosis, molecules are transported to the cell
- 30   surface. Exocytosis facilitates the placement or localization of membrane-bound receptors or other membrane proteins and the secretion of hormones, neurotransmitters, digestive enzymes, and wastes. Endocytosis and exocytosis are fundamental to the function of all

types of cells.

Isolation of intracellular organelles from rat liver has demonstrated the presence of two distinct organelles, the lysosome and the peroxisome. (de Duve, C. (1996) Ann. N.Y. Acad. Sci. 804:1-10.) Lysosomes are the site of degradation of obsolete intracellular  
5 material during autophagy and of extracellular molecules following endocytosis and phagocytosis. They are derived from endosomes, which in turn are formed from budding of the *trans*-Golgi network or from clathrin-coated membrane vesicles invaginating from the plasma membrane.

Protein sorting by transport vesicles, such as the endosome, has important  
10 consequences for a variety of physiological processes including cell surface growth, the biogenesis of distinct intracellular organelles, endocytosis, and the controlled release of hormones and neurotransmitters. (Rothman, J.E. and Wieland, F.T. (1996) Science 272:227-234.) In particular, neurodegenerative disorders and other neuronal pathologies are associated with biochemical flaws during endosomal protein sorting or endosomal  
15 biogenesis. (Mayer R.J. et al. (1996) Adv. Exp. Med. Biol. 389:261-269.)

Peroxisomes are independent organelles and are not members of the secretory pathway family of organelles. They are characterized by a single membrane and a finely granulated matrix and are the site of many peroxide-generating oxidative reactions in the cell. Peroxisomes are unique among eukaryotic organelles in that their size, number, and  
20 enzyme content vary depending upon organism, cell type, and metabolic needs. The majority of peroxisome-associated proteins are membrane-bound or are found proximal to the cytosolic or the luminal side of the peroxisome membrane. (Waterham, H.R. and Cregg, J.M. (1996) BioEssays 19:57-66.)

Genetic defects in peroxisome proteins which result in peroxisomal deficiencies  
25 have been linked to a number of human pathologies, including Zellweger syndrome, rhizomelic chondrodysplasia punctata, X-linked adrenoleukodystrophy, acyl-CoA oxidase deficiency, bifunctional enzyme deficiency, classical Refsum's disease, DHAP alkyl transferase deficiency, and acatalasemia. (Moser, H.W. and Moser, A.B. (1996) Ann. NY Acad. Sci. 804:427-441.) In addition, Gartner, J. et al. (1991; Pediatr. Res. 29:141-146)  
30 found a 22 kDa integral membrane protein associated with lower density peroxisome-like subcellular fractions in patients with Zellweger syndrome.

Normal embryonic development and control of germ cell maturation is modulated

by a number of secretory proteins which interact with their respective membrane-bound receptor. Cell fate during embryonic development is determined, for example, by members of the activin/TGF- $\beta$  superfamily, cadherin(s), IGF-2, and other morphogen(s). In addition, proliferation, maturation, and redifferentiation of germ cell and reproductive  
5 tissues are regulated, for example, by IGF-2, inhibin(s), activin(s), and follistatin(s). (Petraglia, F. (1997) Placenta 18:3-8; and Mather, J.P. et al. (1997) Proc. Soc. Exp. Biol. Med. 215:209-222.)

### **Lymphocyte and Leukocyte Membrane Proteins**

The B-cell response to antigens, which is modulated through receptors, is an  
10 essential component of the normal immune system. Mature B cells recognize foreign antigens through B cell receptors (BCR) and produce specific antibodies which bind the foreign antigens. The antigen/receptor complex is internalized, and the antigen is proteolytically processed. To generate an efficient response to complex antigens, the assistance of the BCR, BCR associated proteins, and T cell is required. A small part of the  
15 antigen remains complexed with major histocompatibility complex-II (MHCII) molecules on the surface of the B cells where the complex can be recognized by T cells. MHCI molecules are on the surface of non-lymphoid tissue and present antigens to macrophages or other lymphoid cell types. T cells recognize and are activated by the MHCI-antigen complex through interactions with the T cell receptor/CD3 complex, a T cell surface  
20 multimeric, multiphenotypic protein located in the plasma membrane. T cells activated by antigen presentation secrete a variety of lymphokines that induce B cell maturation, activate macrophages, and kill target cells.

Leukocytes have a fundamental role in the inflammatory and immune response, and include monocytes/macrophages, mast cells, polymorphonucleoleukocytes, natural killer  
25 cells, neutrophils, eosinophils, basophils, and myeloid precursors. Leukocyte membrane proteins include members of the CD antigens, N-CAM, I-CAM, human leukocyte antigen (HLA) class I and HLA class II gene products, immunoglobulins, immunoglobulin receptors, complement, complement receptors, interferons, interferon receptors, interleukin receptors, and chemokine receptors.

30 Abnormal lymphocyte and leukocyte activity has been associated with, for example, AIDS, immune hypersensitivity, leukemias, leukopenia, systemic lupus, granulomatous disease, and eosinophilia.

## Apoptosis

A variety of ligands and their cellular receptors, enzymes, tumor suppressors, viral gene products, pharmacological agents, and inorganic ions have important positive or negative roles in regulating and implementing the apoptotic destruction of a cell.

- 5 Although some specific components of the apoptotic pathway have been identified and characterized, many interactions between the proteins involved are undefined, leaving major aspects of the pathway unknown.

- A requirement for calcium in apoptosis was previously suggested by studies showing the involvement of calcium levels in DNA cleavage and Fas-mediated cell death.
- 10 (Hewish, D.R. and L.A. Burgoyne (1973) *Biochem. Biophys. Res. Comm.* 52:504-510; Vignaux, F. et al. (1995) *J. Exp. Med.* 181:781-786; and Oshimi, Y. and S. Miyazaki (1995) *J. Immunol.* 154:599-609.) Other studies show that intracellular calcium concentrations increase when apoptosis is triggered in thymocytes by either T cell receptor cross-linking or by glucocorticoids, and cell death can be prevented by blocking this
- 15 increase. (McConkey, D.J. et al. (1989) *J. Immunol.* 143:1801-1806; and McConkey, D.J. et al. (1989) *Arch. Biochem. Biophys.* 269:365-370.)

- The discovery of new human membrane spanning proteins and the polynucleotides encoding it satisfies a need in the art by providing new compositions which are useful in the diagnosis, treatment, and prevention of neoplastic, immunological, and reproductive
- 20 disorders.

## SUMMARY OF THE INVENTION

- The invention features a substantially purified polypeptide, human membrane spanning proteins (MSPs), having an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, and SEQ ID
- 25 NO:5, or fragments thereof.

- The invention further provides a substantially purified variant having at least 90% amino acid identity to the amino acid sequences of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, or to a fragment of any of these sequences. The invention also provides an isolated and purified polynucleotide encoding the polypeptide
- 30 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, or fragments thereof. The invention also includes an isolated and purified polynucleotide variant having at least 90%

polynucleotide sequence identity to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, or fragments thereof.

Additionally, the invention provides an isolated and purified polynucleotide which  
5 hybridizes under stringent conditions to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, or fragments thereof, as well as an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide encoding the polypeptide comprising the amino acid sequence selected  
10 from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, or fragments thereof.

The invention also provides an isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, and SEQ ID NO:11, or fragments  
15 thereof. The invention further provides an isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide sequence comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, or fragment thereof, as well as an isolated and purified polynucleotide having a sequence  
20 which is complementary to the polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, or fragments thereof.

The invention further provides an expression vector containing at least a fragment of the polynucleotide encoding the polypeptide comprising an amino acid sequence  
25 selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, or fragments thereof. In another aspect, the expression vector is contained within a host cell.

The invention also provides a method for producing a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2,  
30 SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, or fragments thereof, the method comprising the steps of: (a) culturing the host cell containing an expression vector containing at least a fragment of a polynucleotide encoding the polypeptide under conditions suitable for the

expression of the polypeptide; and (b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group  
5 consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, or fragments thereof in conjunction with a suitable pharmaceutical carrier.

The invention further includes a purified antibody which binds to a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, or fragments thereof, as well  
10 as a purified agonist and a purified antagonist to the polypeptide.

The invention also provides a method for treating or preventing a neoplastic disorder, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of the polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ  
15 ID NO:4, SEQ ID NO:5, or fragments thereof.

The invention also provides a method for treating or preventing an immunological disorder, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of the polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ  
20 ID NO:4, SEQ ID NO:5, or fragments thereof.

The invention also provides a method for treating or preventing a reproductive disorder, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of the polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ  
25 ID NO:4, SEQ ID NO:5, or fragments thereof.

The invention also provides a method for detecting a polynucleotide encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, or fragments thereof in a biological sample containing nucleic acids, the method comprising the steps  
30 of: (a) hybridizing the complement of the polynucleotide sequence encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, or fragments

thereof to at least one of the nucleic acids of the biological sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide encoding the polypeptide in the biological sample. In one aspect, the nucleic acids of the biological  
5 sample are amplified by the polymerase chain reaction prior to the hybridizing step.

## DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular methodology, protocols, cell  
10 lines, vectors, and reagents described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms  
15 "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the  
20 same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods, devices, and materials are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, vectors, and  
25 methodologies which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

## DEFINITIONS

30 "MSP," as used herein, refers to the amino acid sequences of substantially purified MSP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and preferably the human species, from any source,

whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist," as used herein, refers to a molecule which, when bound to MSP, increases or prolongs the duration of the effect of MSP. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of MSP.

An "allele" or an "allelic sequence," as these terms are used herein, is an alternative form of the gene encoding MSP. Alleles may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to alleles are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding MSP, as described herein, include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide the same MSP or a polypeptide with at least one functional characteristic of MSP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding MSP, and improper or unexpected hybridization to alleles, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding MSP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent MSP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of MSP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine.

The terms "amino acid" or "amino acid sequence," as used herein, refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and



to naturally occurring or synthetic molecules. In this context, "fragments", "immunogenic fragments", or "antigenic fragments" refer to fragments of MSP which are preferably about 5 to about 15 amino acids in length and which retain some biological activity or immunological activity of MSP. Where "amino acid sequence" is recited herein to refer to  
5 an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification," as used herein, relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain  
10 reaction (PCR) technologies well known in the art. (See, e.g., Dieffenbach, C.W. and G.S. Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, pp.1-5.)

The term "antagonist," as it is used herein, refers to a molecule which, when bound to MSP, decreases the amount or the duration of the effect of the biological or  
15 immunological activity of MSP. Antagonists may include proteins, nucleic acids, carbohydrates, antibodies, or any other molecules which decrease the effect of MSP.

As used herein, the term "antibody" refers to intact molecules as well as to fragments thereof, such as Fa, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies that bind MSP polypeptides can be prepared using intact  
20 polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole  
25 limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant," as used herein, refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic  
30 determinants (given regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense," as used herein, refers to any composition containing a nucleic acid sequence which is complementary to a specific nucleic acid sequence. The term "antisense strand" is used in reference to a nucleic acid strand that is complementary to the "sense" strand. Antisense molecules may be produced by any method including  
5 synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" can refer to the antisense strand, and the designation "positive" can refer to the sense strand.

As used herein, the term "biologically active," refers to a protein having structural,  
10 regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic MSP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" or "complementarity," as used herein, refer to the  
15 natural binding of polynucleotides under permissive salt and temperature conditions by base pairing. For example, the sequence "A-G-T" binds to the complementary sequence "T-C-A." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of  
20 complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" or a "composition  
25 comprising a given amino acid sequence," as these terms are used herein, refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation, an aqueous solution, or a sterile composition. Compositions comprising polynucleotide sequences encoding MSP or fragments of MSP may be employed as hybridization probes. The probes may be stored in  
30 freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., SDS), and other components (e.g., Denhardt's solution, dry milk,

salmon sperm DNA, etc.).

The phrase "consensus sequence," as used herein, refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using XL-PCR™ (Perkin Elmer, Norwalk, CT) in the 5' and/or the 3' direction, and resequenced, or which has been  
5 assembled from the overlapping sequences of more than one Incyte Clone using a computer program for fragment assembly, such as the GELVIEW™ Fragment Assembly system (GCG, Madison, WI). Some sequences have been both extended and assembled to produce the consensus sequence .

As used herein, the term "correlates with expression of a polynucleotide" indicates  
10 that the detection of the presence of nucleic acids, the same or related to a nucleic acid sequence encoding MSP, by northern analysis is indicative of the presence of nucleic acids encoding MSP in a sample, and thereby correlates with expression of the transcript from the polynucleotide encoding MSP.

A "deletion," as the term is used herein, refers to a change in the amino acid or  
15 nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative," as used herein, refers to the chemical modification of MSP, of a polynucleotide sequence encoding MSP, or of a polynucleotide sequence complementary to a polynucleotide sequence encoding MSP. Chemical modifications of a  
20 polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it  
25 was derived.

The term "homology," as used herein, refers to a degree of complementarity. There may be partial homology or complete homology. The word "identity" may substitute for the word "homology." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred  
30 to as "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization, and the like) under conditions of

reduced stringency. A substantially homologous sequence or hybridization probe will compete for and inhibit the binding of a completely homologous sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% homology or identity). In the absence of non-specific binding, the substantially homologous sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" or "% identity" refer to the percentage of sequence similarity found in a comparison of two or more amino acid or nucleic acid sequences. Percent identity can be determined electronically, e.g., by using the MegAlign program (DNASTAR, Inc., Madison WI). The MegAlign program can create alignments between two or more sequences according to different methods, e.g., the Clustal method. (See, e.g., Higgins, D.G. and P. M. Sharp (1988) *Gene* 73:237-244.) The Clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. The percentage similarity between two amino acid sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no homology between the two amino acid sequences are not included in determining percentage similarity. Percent identity between nucleic acid sequences can also be counted or calculated by other methods known in the art, e.g., the Jotun Hein method. (See, e.g., Hein, J. (1990) *Methods Enzymol.* 183:626-645.) Identity between sequences can also be determined by other methods known in the art, e.g., by varying hybridization conditions.

"Human artificial chromosomes" (HACs), as described herein, are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance. (See, e.g., Harrington, J.J. et al. (1997) *Nat Genet.* 15:345-355.)

The term "humanized antibody," as used herein, refers to antibody molecules in

which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

“Hybridization,” as the term is used herein, refers to any process by which a strand  
5 of nucleic acid binds with a complementary strand through base pairing.

As used herein, the term “hybridization complex” as used herein, refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g.,  $C_0t$  or  $R_0t$  analysis) or formed between one nucleic acid sequence present  
10 in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words “insertion” or “addition,” as used herein, refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or  
15 nucleotides, respectively, to the sequence found in the naturally occurring molecule.

“Immune response” can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

20 The term “microarray,” as used herein, refers to an arrangement of distinct polynucleotides arrayed on a substrate, e.g., paper, nylon or any other type of membrane, filter, chip, glass slide, or any other suitable solid support.

The terms “element” or “array element” as used herein in a microarray context, refer to hybridizable polynucleotides arranged on the surface of a substrate.

25 The term “modulate,” as it appears herein, refers to a change in the activity of MSP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of MSP.

The phrases “nucleic acid” or “nucleic acid sequence,” as used herein, refer to an  
30 oligonucleotide, nucleotide, polynucleotide, or any fragment thereof, to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-

like or RNA-like material. In this context, "fragments" refers to those nucleic acid sequences which are greater than about 60 nucleotides in length, and most preferably are at least about 100 nucleotides, at least about 1000 nucleotides, or at least about 10,000 nucleotides in length.

5           The terms "operably associated" or "operably linked," as used herein, refer to functionally related nucleic acid sequences. A promoter is operably associated or operably linked with a coding sequence if the promoter controls the transcription of the encoded polypeptide. While operably associated or operably linked nucleic acid sequences can be contiguous and in reading frame, certain genetic elements, e.g., repressor genes, are not  
10           contiguously linked to the encoded polypeptide but still bind to operator sequences that control expression of the polypeptide.

          The term "oligonucleotide," as used herein, refers to a nucleic acid sequence of at least about 6 nucleotides to 60 nucleotides, preferably about 15 to 30 nucleotides, and most preferably about 20 to 25 nucleotides, which can be used in PCR amplification or in  
15           a hybridization assay or microarray. As used herein, the term "oligonucleotide" is substantially equivalent to the terms "amplimer," "primer," "oligomer," and "probe," as these terms are commonly defined in the art.

          "Peptide nucleic acid" (PNA), as used herein, refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in  
20           length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA and RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell. (See, e.g., Nielsen, P.E. et al. (1993) Anticancer Drug Des. 8:53-63.)

25           The term "sample," as used herein, is used in its broadest sense. A biological sample suspected of containing nucleic acids encoding MSP, or fragments thereof, or MSP itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a solid support; a tissue; a tissue print; etc.

30           As used herein, the terms "specific binding" or "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, or an antagonist. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the

antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

5 As used herein, the term "stringent conditions" refers to conditions which permit hybridization between polynucleotide sequences and the claimed polynucleotide sequences. Suitably stringent conditions can be defined by, for example, the concentrations of salt or formamide in the prehybridization and hybridization solutions, or by the hybridization temperature, and are well known in the art. In particular, stringency  
10 can be increased by reducing the concentration of salt, increasing the concentration of formamide, or raising the hybridization temperature.

For example, hybridization under high stringency conditions could occur in about 50% formamide at about 37°C to 42°C. Hybridization could occur under reduced stringency conditions in about 35% to 25% formamide at about 30°C to 35°C. In  
15 particular, hybridization could occur under high stringency conditions at 42°C in 50% formamide, 5X SSPE, 0.3% SDS, and 200 µg/ml sheared and denatured salmon sperm DNA. Hybridization could occur under reduced stringency conditions as described above, but in 35% formamide at a reduced temperature of 35°C. The temperature range corresponding to a particular level of stringency can be further narrowed by calculating the  
20 purine to pyrimidine ratio of the nucleic acid of interest and adjusting the temperature accordingly. Variations on the above ranges and conditions are well known in the art.

The term "substantially purified," as used herein, refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably  
25 about 90% free from other components with which they are naturally associated.

A "substitution," as used herein, refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Transformation," as defined herein, describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial  
30 conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host

cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host  
5 chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "variant" of MSP, as used herein, refers to an amino acid sequence that is altered by one or more amino acids. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (e.g., replacement of  
10 leucine with isoleucine). More rarely, a variant may have "nonconservative" changes (e.g., replacement of glycine with tryptophan). Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for  
15 example, DNASTAR software.

#### THE INVENTION

The invention is based on the discovery of new human membrane spanning proteins, collectively referred to as MSP and individually as MSP-1, MSP-2, MSP-3,  
20 MSP-4, MSP-5, and MSP-6; the polynucleotides encoding MSP (SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, and SEQ ID NO:11); and the use of these compositions for the diagnosis, treatment, or prevention of neoplastic, immunological, and reproductive disorders. Table 1 shows the sequence identification numbers, Incyte Clone identification number, cDNA library, public database (PD)  
25 homolog sequence identifier and homolog species description for each of the human membrane spanning proteins disclosed herein.

Nucleic acids encoding the MSP-1 of the present invention were first identified in Incyte Clone 77138 from the synovial membrane tissue cDNA library (SYNORAB01) using a computer search for amino acid sequence alignments. A consensus sequence,  
30 SEQ ID NO:6, was derived from Incyte Clones 77138 (SYNORAB01), 3576995 (BRONNOT01), 1995355 (BRSTTUT03), and 1260677 (SYNORAT05).

In one embodiment, the invention encompasses a polypeptide comprising the



amino acid sequence of SEQ ID NO:1. MSP-1 is 238 amino acids in length and has two potential N glycosylation sites at residues N71 and N72; one potential cAMP- and cGMP-dependent protein kinase phosphorylation site at residue S233; three potential casein kinase II phosphorylation sites at residues T111, T131, and T195; four potential  
5 protein kinase C phosphorylation sites at residues T10, S44, T186, and S233; and the PMP-22/EMP/MP20 membrane protein family signature from about residue V205 to about residue M232; a proline-rich region from about P26 to about P41; and the G-protein alpha subunit signature from about R61 to about T76. MSP-1 shares 63% identity with the dog mucin-type gp40 protein (g1628360). In addition, the PMP-  
10 22/EMP/MP20 membrane protein family signature is conserved between these molecules. A fragment of SEQ ID NO:6 from about nucleotide 379 to about nucleotide 405 is useful for designing oligonucleotides or for use as a hybridization probe. Northern analysis shows the expression of this sequence in reproductive, nervous, and connective tissue cDNA libraries. Approximately 56% of these libraries are associated  
15 with neoplastic disorders and 23% with immune response.

Nucleic acids encoding the MSP-2 of the present invention were first identified in Incyte Clone 1381884 from the brain cDNA library (BRAITUT08) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:7, was derived from Incyte Clones 1381884 (BRAITUT08), 1862104 (PROSNOT19), 79810  
20 (SYNORAB01), 2372040 (ADRENOT07), 1811401 (PROSTUT12), 1811401 (PROSTUT12), and 1927265 (BRSTNOT02).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:2. MSP-2 is 83 amino acids in length and has one potential N glycosylation site at residue N80; one potential casein kinase II  
25 phosphorylation site at residue T71; and the inhibin beta A chain signature from about residue E39 to about residue K60. MSP-2 shares 51% identity with human thyroid receptor interactor (g927071). In addition, the C-terminal glycine residue spacings G41-{5N}-G47-{2N}-G50-{8N}-G59-{2N}-G62-{5N}-G69, where N is any amino acid, are conserved between these molecules. A fragment of SEQ ID NO:7 from about  
30 nucleotide 355 to about nucleotide 375 is useful for designing oligonucleotides or for use as a hybridization probe. Northern analysis shows the expression of this sequence in reproductive, cardiovascular, nervous, and gastrointestinal cDNA libraries.

TABLE 1

Protein	Nucleotide	Clone ID	Library	PD Homolog	Homolog species
SEQ ID NO:1	SEQ ID NO:6	77138	SYNORAB01	g1628360	Canis familiaris
SEQ ID NO:2	SEQ ID NO:7	1381884	BRAITUT08	g927071	Homo sapiens
SEQ ID NO:3	SEQ ID NO:8	1457779	COLNFET02	g765256	Homo sapiens
SEQ ID NO:4	SEQ ID NO:9	1481261	CORPNOT02	g177900	Homo sapiens
SEQ ID NO:5	SEQ ID NO:10	1794154	PROSTUT05	WO9640907-A1	Homo sapiens
	SEQ ID NO:11	1737775	COLNNOT22	g1184066	Bos taurus

Approximately 43 % of these libraries are associated with neoplastic disorders and 21 % with immune response.

Nucleic acids encoding the MSP-3 of the present invention were first identified in Incyte Clone 1457779 from the fetal colon cDNA library (COLNFET02) using a  
5 computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:8, was derived from Incyte Clones 1457779 and 1457572 (COLNFET02), 2256441 (OVARTUT01), and the shotgun sequences SAGA01244, SAGA00103, SAGA01339, SAGA00583, and SAGA01282.

In one embodiment, the invention encompasses a polypeptide comprising the  
10 amino acid sequence of SEQ ID NO:3. MSP-3 is 495 amino acids in length and has seven potential N glycosylation sites at residues N67, N135, N304, N325, N363, N374, and N447; one potential cAMP- and cGMP-dependent protein kinase phosphorylation site at residue S136; four potential casein kinase II phosphorylation sites at residues S4, S153, T226, and T376; three potential protein kinase C phosphorylation sites at residues  
15 S4, S132, and T365; one potential tyrosine kinase phosphorylation site at residue Y392; a potential signal peptide sequence from about residue M1 to about G21; ten cystein residues from about C78 to about C403; and four GDA1/CD39 family transmembrane signatures: TM1 extends from about F43 to about Y59; TM2, from about T118 to about L131; TM3, from about L162 to about L183; and TM4, from about G202 to about  
20 F215. MSP-3 shares 46% identity with human CD39 homolog (g765256). The cysteines at C78, C102, C246, C273, C292, C316, C329, C335, C381, and C403 are conserved across both molecules. In addition, the hydrophobic transmembrane domains are conserved between these molecules. A fragment of SEQ ID NO:8 from about nucleotide 1543 to about nucleotide 1566 is useful for designing oligonucleotides or for  
25 use as a hybridization probe. Northern analysis shows the expression of this sequence in gastrointestinal and ovarian cDNA libraries. Approximately 71 % of these libraries are associated with neoplastic disorders and 14 % with immune response.

Nucleic acids encoding the MSP-4 of the present invention were first identified in Incyte Clone 1481261 from the corpus callosum cDNA library (CORPNOT02) using  
30 a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:9, was derived from Incyte Clones 1481261 and 1476703 (CORPNOT02), and 2963217 (SCORNOT04).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:4. MSP-4 is 156 amino acids in length and has three potential casein kinase II phosphorylation sites at residues S3, T114, and S151; three potential protein kinase C phosphorylation sites at residues S3, S32, and T114; 5 and a potential GPR1/FUN34/yaaH family transmembrane domain from about residue F95 to about residue I141. MSP-4 shares 25 % identity with human differentiation-dependent intestinal A4 transmembrane protein (g177900). In addition, the hydrophobic transmembrane domain and one potential protein kinase C phosphorylation site are conserved between these molecules. A fragment of SEQ ID NO:9 from about 10 nucleotide 373 to about nucleotide 396 is useful for designing oligonucleotides or for use as a hybridization probe. Northern analysis shows the expression of this sequence in brain, spinal cord, reproductive, muscle, and Alzheimer's disease brain tissue cDNA libraries. Approximately 64% of these libraries are associated with neoplastic disorders and 19% with immune response.

15 Nucleic acids encoding the MSP-5 of the present invention were first identified in Incyte Clone 1794154 from the prostate cDNA library (PROSTUT05) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:10, was derived from Incyte Clones 1794154 (PROSTUT05), 3116657 (LUNGTUT13), 1726333 (PROSNOT14), 3217158 (TESTNOT07), 1435295 20 (PANCNOT08), 2530829 (GBLANOT02), 728143 (SYNOOAT01), 1539910 (SINTTUT01), 1467860 (PANCTUT02), 1484333 (CORPNOT02), 1927469 (BRSTNOT02), 1516116 (PANCTUT01), 1571055 (UTRSNOT05), 1965064 (BRSTNOT04), and 2683922 (LUNGNOT23).

In one embodiment, the invention encompasses a polypeptide comprising the 25 amino acid sequence of SEQ ID NO:5. MSP-5 is 950 amino acids in length and has three potential N glycosylation site at residues N667 N668 N835; six potential cAMP- and cGMP-dependent protein kinase phosphorylation sites at residues S79, T204, S409, S434, S658, and S767; seventeen potential casein kinase II phosphorylation sites at residues S64, T65, S113, S284, T398, S409, S418, T437, T449, S518, S576, T591, 30 S611, T708, T717, S735, and S838; twenty five potential protein kinase C phosphorylation sites at residues T2, T28, S45, T55, S113, S124, S308, T347, T365, T384, S393, T404, S418, T429, T430, S454, S537, S565, S572, S583, T600, S607,

T654, T670, and T732; one potential tyrosine kinase phosphorylation site at residue Y918; and a potential signal peptide sequence from about M1 to P23. MSP-5 shares 28% identity with human type I, p80 IL-1-receptor intracellular domain ligand (WO9640907-A1). A fragment of SEQ ID NO:10 from about nucleotide 3220 to about  
5 nucleotide 3243 is useful for designing oligonucleotides or for use as a hybridization probe. Northern analysis shows the expression of this sequence in gastrointestinal, male reproductive, and muscle cDNA libraries. Approximately 54% of these libraries are associated with neoplastic disorders and 22% with immune response.

Nucleic acids encoding the MSP-6 of the present invention were first identified  
10 in Incyte Clone 1737775 from the colon cDNA library (COLNNOT22) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:11, was derived from Incyte Clones 1737775 and 1734248 (COLNNOT22), 1578253 (DUODNOT01), 2214423 (SINTFET03), 608819 (COLNNOT01), 1629002 and 1626949 (COLNPOT01), 1498226 and 1429362 (SINTBST01), 2925341  
15 (SININOT04), and the shotgun sequences SAEC10115, SAEA00623, and SAEA00834.

In one embodiment, a fragment of SEQ ID NO:11 from about nucleotide 1447 to about nucleotide 1485 is useful for designing oligonucleotides or for use as a hybridization probe. Northern analysis shows the expression of this sequence in gastrointestinal, female reproductive, and cardiovascular cDNA libraries.  
20 Approximately 32% of these libraries are associated with neoplastic disorders and 56% with immune response.

The invention also encompasses MSP variants. A preferred MSP variant is one which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the MSP amino acid sequence, and which  
25 contains at least one functional or structural characteristic of MSP.

The invention also encompasses polynucleotides which encode MSP. Accordingly, any nucleic acid sequence which encodes the amino acid sequence of MSP can be used to produce recombinant molecules which express MSP. In a particular embodiment, the invention encompasses a polynucleotide consisting of a nucleic acid  
30 sequence selected from the group consisting of SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, and SEQ ID NO:11.

The invention also encompasses a variant of a polynucleotide sequence encoding

MSP. In particular, such a variant polynucleotide sequence will have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding MSP. A particular aspect of the invention encompasses a variant of a nucleic acid sequence selected from the group  
5 consisting of SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, and SEQ ID NO:11 which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide consisting of a nucleic acid sequence selected from the group consisting of SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, and SEQ ID NO:11. Any one of the  
10 polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of MSP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding MSP, some bearing minimal homology to the polynucleotide sequences of any known and naturally occurring  
15 gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring MSP, and all such variations are to be considered as being specifically disclosed.

20 Although nucleotide sequences which encode MSP and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring MSP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding MSP or its derivatives possessing a substantially different codon usage. Codons may be selected to increase the rate at which expression of the  
25 peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding MSP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced  
30 from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode MSP

and MSP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents that are well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence  
5 encoding MSP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed nucleotide sequences, and in particular, those shown in SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, and fragments thereof, under various conditions of stringency. (See, e.g., Wahl, G.M. and  
10 S.L. Berger (1987) *Methods Enzymol.* 152:399-407; and Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.)

Methods for DNA sequencing are well known and generally available in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase® (US  
15 Biochemical Corp., Cleveland, OH), Taq polymerase (Perkin Elmer), thermostable T7 polymerase (Amersham, Chicago, IL), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE Amplification System (GIBCO/BRL, Gaithersburg, MD). Preferably, the process is automated with machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno, NV), Peltier Thermal Cycler  
20 (PTC200; MJ Research, Watertown, MA) and the ABI Catalyst and 373 and 377 DNA Sequencers (Perkin Elmer).

The nucleic acid sequences encoding MSP may be extended utilizing a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which  
25 may be employed, restriction-site PCR, uses universal primers to retrieve unknown sequence adjacent to a known locus. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.) In particular, genomic DNA is first amplified in the presence of a primer which is complementary to a linker sequence within the vector and a primer specific to a region of the nucleotide sequence. The amplified sequences are then subjected to a second  
30 round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR may also be used to amplify or extend sequences using divergent primers based on a known region. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186.) The primers may be designed using commercially available software such as OLIGO 4.06 Primer Analysis software (National Biosciences Inc., Plymouth, MN) or  
5 another appropriate program to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to 72°C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

10 Another method which may be used is capture PCR, which involves PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) *PCR Methods Applic.* 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to place an engineered double-stranded sequence into an unknown fragment  
15 of the DNA molecule before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) *Nucleic Acids Res.* 19:3055-3060.) Additionally, one may use PCR, nested primers, and PromoterFinder™ libraries to walk genomic DNA (Clontech, Palo Alto, CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

20 When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable in that they will include more sequences which contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension  
25 of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different fluorescent dyes (one for each nucleotide) which are laser  
30 activated, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., Genotyper™ and Sequence Navigator™, Perkin Elmer), and the entire process from



loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for the sequencing of small pieces of DNA which might be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments  
5 thereof which encode MSP may be used in recombinant DNA molecules to direct expression of MSP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced, and these sequences may be used to clone and express MSP.

10 As will be understood by those of skill in the art, it may be advantageous to produce MSP-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce an RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated  
15 from the naturally occurring sequence.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter MSP-encoding sequences for a variety of reasons including, but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR  
20 reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, introduce mutations, and so forth.

In another embodiment of the invention, natural, modified, or recombinant nucleic  
25 acid sequences encoding MSP may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of MSP activity, it may be useful to encode a chimeric MSP protein that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between the MSP encoding sequence and the heterologous protein sequence, so  
30 that MSP may be cleaved and purified away from the heterologous moiety.

In another embodiment, sequences encoding MSP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al.

(1980) Nucl. Acids Res. Symp. Ser. 215-223, and Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser. 225-232.) Alternatively, the protein itself may be produced using chemical methods to synthesize the amino acid sequence of MSP, or a fragment thereof. For example, peptide synthesis can be performed using various solid-phase techniques. (See, 5 e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A Peptide Synthesizer (Perkin Elmer). Additionally, the amino acid sequence of MSP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

10 The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1983) Proteins, Structures and Molecular Properties, WH Freeman and Co., New York, NY.)

15 In order to express a biologically active MSP, the nucleotide sequences encoding MSP or derivatives thereof may be inserted into appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence.

Methods which are well known to those skilled in the art may be used to construct 20 expression vectors containing sequences encoding MSP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, ch. 4, 8, and 16-17; and Ausubel, F.M. et al. (1995, and periodic 25 supplements) Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding MSP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression 30 vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus (CaMV) or tobacco mosaic virus

(TMV)) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems.

The invention is not limited by the host cell employed.

The "control elements" or "regulatory sequences" are those non-translated regions, e.g., enhancers, promoters, and 5' and 3' untranslated regions, of the vector and polynucleotide sequences encoding MSP which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters, e.g., hybrid lacZ promoter of the Bluescript® phagemid (Stratagene, La Jolla, CA) or pSport1™ plasmid (GIBCO/BRL), may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO, and storage protein genes) or from plant viruses (e.g., viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding MSP, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for MSP. For example, when large quantities of MSP are needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, multifunctional E. coli cloning and expression vectors such as Bluescript® (Stratagene), in which the sequence encoding MSP may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of  $\beta$ -galactosidase so that a hybrid protein is produced, and pIN vectors. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) pGEX vectors (Pharmacia Biotech, Uppsala, Sweden) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the

cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast Saccharomyces cerevisiae, a number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH, may be used. (See, e.g., Ausubel, supra; and Grant et al. (1987) *Methods Enzymol.* 153:516-544.)

5 In cases where plant expression vectors are used, the expression of sequences encoding MSP may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV. (Takamatsu, N. (1987) *EMBO J.* 6:307-311.) Alternatively, plant promoters such as the small subunit of RUBISCO or  
10 heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews. (See, e.g., Hobbs, S. or Murry, L.E.  
15 in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, NY; pp. 191-196.)

An insect system may also be used to express MSP. For example, in one such system, Autographa californica nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in Spodoptera frugiperda cells or in Trichoplusia larvae. The  
20 sequences encoding MSP may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of sequences encoding MSP will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, S. frugiperda cells or Trichoplusia larvae in which MSP may be  
25 expressed. (See, e.g., Engelhard, E.K. et al. (1994) *Proc. Nat. Acad. Sci.* 91:3224-3227.)

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding MSP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region  
30 of the viral genome may be used to obtain a viable virus which is capable of expressing MSP in infected host cells. (See, e.g., Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci.* 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus

(RSV) enhancer, may be used to increase expression in mammalian host cells.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, 5 polycationic amino polymers, or vesicles) for therapeutic purposes.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding MSP. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding MSP and its initiation codon and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional 10 or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and 15 synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such 20 modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding, and/or function. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, 25 HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC, Bethesda, MD) and may be chosen to ensure the correct modification and processing of the foreign protein.

For long term, high yield production of recombinant proteins, stable expression is preferred. For example, cell lines capable of stably expressing MSP can be transformed 30 using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2

days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase genes and adenine phosphoribosyltransferase genes, which can be employed in *tk* or *apr*<sup>r</sup> cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; and Lowy, I. et al. (1980) Cell 22:817-823) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *npt* confers resistance to the aminoglycosides neomycin and G-418; and *als* or *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14; and Murry, supra.) Additional selectable genes have been described, e.g., *trpB*, which allows cells to utilize indole in place of tryptophan, or *hisD*, which allows cells to utilize histinol in place of histidine. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-8051.) Visible markers, e.g., anthocyanins,  $\beta$  glucuronidase and its substrate GUS, luciferase and its substrate luciferin may be used. Green fluorescent proteins (GFP) (Clontech, Palo Alto, CA) can also be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. et al. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding MSP is inserted within a marker gene sequence, transformed cells containing sequences encoding MSP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding MSP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells which contain the nucleic acid sequence encoding MSP

and express MSP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid  
5 or protein sequences.

The presence of polynucleotide sequences encoding MSP can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes or fragments or fragments of polynucleotides encoding MSP. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the sequences encoding MSP to  
10 detect transformants containing DNA or RNA encoding MSP.

A variety of protocols for detecting and measuring the expression of MSP, using either polyclonal or monoclonal antibodies specific for the protein, are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site,  
15 monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on MSP is preferred, but a competitive binding assay may be employed. These and other assays are well described in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN, Section IV; and Maddox, D.E. et al. (1983) J. Exp. Med. 158:1211-1216).

20 A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding MSP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding  
25 MSP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Pharmacia & Upjohn  
30 (Kalamazoo, MI), Promega (Madison, WI), and U.S. Biochemical Corp. (Cleveland, OH). Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as

substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding MSP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or contained intracellularly

5 depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode MSP may be designed to contain signal sequences which direct secretion of MSP through a prokaryotic or eukaryotic cell membrane. Other constructions may be used to join sequences encoding MSP to nucleotide sequences encoding a polypeptide domain which will facilitate

10 purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, WA). The inclusion of cleavable linker

15 sequences, such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, CA), between the purification domain and the MSP encoding sequence may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing MSP and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification

20 on immobilized metal ion affinity chromatography (IMAC). (See, e.g., Porath, J. et al. (1992) Prot. Exp. Purif. 3: 263-281.) The enterokinase cleavage site provides a means for purifying MSP from the fusion protein. (See, e.g., Kroll, D.J. et al. (1993) DNA Cell Biol. 12:441-453.)

Fragments of MSP may be produced not only by recombinant production, but also

25 by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, T.E. (1984) Protein: Structures and Molecular Properties, pp. 55-60, W.H. Freeman and Co., New York, NY.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Various fragments of MSP may be

30 synthesized separately and then combined to produce the full length molecule.

## THERAPEUTICS



Chemical and structural homology exists among the human membrane spanning proteins of the invention. In addition, the expression of MSP is closely associated with cell proliferation, and with tissues associated with cancer, the immune response, and in reproductive tissues. Therefore, MSP appears to play a role in cancer, immunological, and  
5 reproductive disorders, in particular where increased activity or synthesis appears to be associated with these disorders.

In one embodiment, antagonists which decrease the expression or activity of MSP may be administered to a subject to treat or prevent a neoplastic disorder such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, and  
10 teratocarcinoma. Such cancers include, but are not limited to, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. In one aspect, antibodies which specifically bind MSP may be used directly as an antagonist or indirectly as a  
15 targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express MSP.

In another embodiment, a vector expressing the complement of the polynucleotide encoding MSP may be administered to a subject to treat or prevent a neoplastic disorder including, but not limited to, those listed above.

20 In yet another embodiment, antagonists which decrease the activity of MSP may be administered to a subject to treat or prevent an immunological disorder. Such immunological disorders may be associated with conditions such as AIDS, Addison's disease, adult respiratory distress syndrome, allergies, anemia, asthma, atherosclerosis, bronchitis, cholecystitis, Crohn's disease, ulcerative colitis, atopic dermatitis,  
25 dermatomyositis, diabetes mellitus, emphysema, atrophic gastritis, glomerulonephritis, gout, Graves' disease, hypereosinophilia, irritable bowel syndrome, lupus erythematosus, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis, scleroderma, Sjögren's syndrome, and autoimmune thyroiditis; complications of cancer, hemodialysis,  
30 extracorporeal circulation; viral, bacterial, fungal, parasitic, protozoal, and helminthic infections; and trauma. In one aspect, antibodies which specifically bind MSP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for

bringing a pharmaceutical agent to cells or tissue which express MSP.

In another embodiment, a vector expressing the complement of the polynucleotide encoding MSP may be administered to a subject to treat or prevent an immunological disorder including, but not limited to, those listed above.

5 In a further embodiment, an antagonist of MSP may be administered to a subject to treat or prevent a reproductive disorder. Such a reproductive disorder may include, but is not limited to, disorders of prolactin production; infertility, including tubal disease, ovulatory defects, and endometriosis; disruptions of the estrous cycle, disruptions of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome,  
10 endometrial and ovarian tumors, autoimmune disorders, ectopic pregnancy, and teratogenesis; cancer of the breast, uterine fibroids, fibrocystic breast disease, galactorrhea; disruptions of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, carcinoma of the male breast and gynecomastia. In one aspect, an antibody which specifically binds MSP  
15 may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express MSP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding MSP may be administered to a subject to treat or prevent a reproductive disorder including, but not limited to, those described above.

20 In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act  
25 synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of MSP may be produced using methods which are generally known in the art. In particular, purified MSP may be used to produce antibodies or to screen  
30 libraries of pharmaceutical agents to identify those which specifically bind MSP.

Antibodies to MSP may also be generated using methods that are well known in the art.

Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and

single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice,  
5 humans, and others may be immunized by injection with MSP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil  
10 emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to MSP have an amino acid sequence consisting of at least about 5 amino acids, and, more preferably, of at least about 10 amino acids. It is also preferable that these  
15 oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of MSP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to MSP may be prepared using any technique which  
20 provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. 80:2026-2030; and Cole, S.P. et al. (1984) Mol.  
25 Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. 81:6851-6855; Neuberger, M.S. et al.  
30 (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.)

Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce MSP-specific single chain antibodies.

Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton D.R. (1991) Proc. Natl. Acad. Sci. 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the  
5 lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. 86: 3833-3837; and Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for MSP may also be generated. For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments  
10 produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

15 Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between MSP and its specific antibody. A two-site, monoclonal-based  
20 immunoassay utilizing monoclonal antibodies reactive to two non-interfering MSP epitopes is preferred, but a competitive binding assay may also be employed. (Maddox, supra.)

In another embodiment of the invention, the polynucleotides encoding MSP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the  
25 complement of the polynucleotide encoding MSP may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding MSP. Thus, complementary molecules or fragments may be used to modulate MSP activity, or to achieve regulation of gene function. Such technology is now well known in the art, and  
30 sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding MSP.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia

viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors which will express nucleic acid sequences complementary to the polynucleotides of the gene encoding MSP. (See, e.g.,  
5 Sambrook, supra; and Ausubel, supra.)

Genes encoding MSP can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding MSP. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such  
10 vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by  
15 designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding MSP. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the  
20 ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to  
25 block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme  
30 molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding MSP.

Specific ribozyme cleavage sites within any potential RNA target are initially

identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may  
5 render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules.  
10 These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding MSP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that  
15 synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather  
20 than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

25 Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in  
30 the art. (See, e.g., Goldman, C.K. et al. (1997) Nature Biotechnology 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses,

rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical  
5 compositions may consist of MSP, antibodies to MSP, and mimetics, agonists, antagonists, or inhibitors of MSP. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a  
10 patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

15 In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing  
20 Co., Easton, PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for  
25 ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as  
30 sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including

arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as  
5 concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

10 Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or  
15 suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiologically buffered saline. Aqueous injection  
20 suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino  
25 polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the  
30 art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving,



granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, 5 malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

10 After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of MSP, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions 15 wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells or in animal models such as mice, rats, 20 rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example MSP or fragments thereof, antibodies of MSP, and agonists, antagonists or 25 inhibitors of MSP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED<sub>50</sub> (the dose therapeutically effective in 50% of the population) or LD<sub>50</sub> (the dose lethal to 50% of the population) statistics. The dose ratio of therapeutic to toxic effects is the therapeutic index, and it can 30 be expressed as the ED<sub>50</sub>/LD<sub>50</sub> ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in

such compositions is preferably within a range of circulating concentrations that includes the ED<sub>50</sub> with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu$ g to 100,000  $\mu$ g, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## 20 **DIAGNOSTICS**

In another embodiment, antibodies which specifically bind MSP may be used for the diagnosis of disorders characterized by expression of MSP, or in assays to monitor patients being treated with MSP or agonists, antagonists, or inhibitors of MSP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for MSP include methods which utilize the antibody and a label to detect MSP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

30 A variety of protocols for measuring MSP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of MSP expression. Normal or standard values for MSP expression are established by combining

body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to MSP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, preferably by photometric means. Quantities of MSP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding MSP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of MSP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of MSP, and to monitor regulation of MSP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding MSP or closely related molecules may be used to identify nucleic acid sequences which encode MSP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low), will determine whether the probe identifies only naturally occurring sequences encoding MSP, alleles, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably have at least 50% sequence identity to any of the MSP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and derived from the nucleotide sequence of SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, or from genomic sequences including promoters, enhancers, and introns of the MSP gene.

Means for producing specific hybridization probes for DNAs encoding MSP include the cloning of polynucleotide sequences encoding MSP or MSP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides.

Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding MSP may be used for the diagnosis of a  
5 disorder associated with expression of MSP. Examples of such a disorder include, but are not limited to, a neoplastic disorder such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, and teratocarcinoma. Such cancers include, but are not limited to, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary,  
10 pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an immunological disorder such as AIDS, Addison's disease, adult respiratory distress syndrome, allergies, anemia, asthma, atherosclerosis, bronchitis, cholecystitis, Crohn's disease, ulcerative colitis, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, atrophic gastritis, glomerulonephritis, gout, Graves' disease,  
15 hypereosinophilia, irritable bowel syndrome, lupus erythematosus, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis, scleroderma, Sjögren's syndrome, and autoimmune thyroiditis; complications of cancer, hemodialysis, extracorporeal circulation; viral, bacterial, fungal, parasitic, protozoal, and helminthic infections; and trauma; a  
20 reproductive disorder such as, disorders of prolactin production; infertility, including tubal disease, ovulatory defects, and endometriosis; disruptions of the estrous cycle, disruptions of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, autoimmune disorders, ectopic pregnancy, and teratogenesis; cancer of the breast, uterine fibroids, fibrocystic breast disease, galactorrhea;  
25 disruptions of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, carcinoma of the male breast and gynecomastia. The polynucleotide sequences encoding MSP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and ELISA assays; and in microarrays utilizing fluids or  
30 tissues from patients to detect altered MSP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding MSP may be useful in

assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding MSP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the  
5 signal is quantitated and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding MSP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to  
10 monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of MSP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding MSP, under  
15 conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard  
20 values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of  
25 treatment over a period ranging from several days to months.

With respect to cancer, the presence of a relatively high amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health  
30 professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences

encoding MSP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding MSP, or a fragment of a polynucleotide complementary to the polynucleotide encoding MSP, and will be employed under  
5 optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of MSP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and  
10 interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; and Duplaa, C. et al. (1993) Anal. Biochem. 229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

15 In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a  
20 disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler et al. (1995) PCT application  
25 WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding MSP may be used to generate hybridization probes useful in mapping the naturally occurring  
30 genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial

chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Price, C.M. (1993) *Blood Rev.* 7:127-134; and Trask, B.J. (1991) *Trends Genet.* 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical  
5 chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, R.A. (ed.) Molecular Biology and Biotechnology, VCH Publishers New York, NY, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) site. Correlation between the location of the gene encoding MSP on a physical chromosomal map and a specific  
10 disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used  
15 for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery  
20 techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., AT to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) *Nature* 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation,  
25 inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, MSP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located  
30 intracellularly. The formation of binding complexes between MSP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of

compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with MSP, or fragments thereof, and  
5 washed. Bound MSP is then detected by methods well known in the art. Purified MSP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which  
10 neutralizing antibodies capable of binding MSP specifically compete with a test compound for binding MSP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with MSP.

In additional embodiments, the nucleotide sequences which encode MSP may be used in any molecular biology techniques that have yet to be developed, provided the new  
15 techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention.

20

## EXAMPLES

### I. COLNNOT22 cDNA Library Construction

The COLNNOT22 library was constructed from microscopically normal colon tissue excised from a 56-year-old Caucasian female during a resection of the small  
25 intestine. The patient was diagnosed with Crohn's disease involving the ileum and ileal-colonic anastomosis. Patient history included a cholecystectomy and breast lesions. Family history included atherosclerosis in a grandparent and functional disorder of the intestine in the patient's mother.

The frozen tissue was homogenized and lysed using a Brinkmann Homogenizer  
30 Polytron PT-3000 (Brinkmann Instruments, Westbury, NJ) in guanidinium isothiocyanate solution. The lysate was centrifuged over a 5.7 M CsCl cushion using an Beckman SW28 rotor in a Beckman L8-70M Ultracentrifuge (Beckman Instruments) for 18 hours at 25,000



rpm at ambient temperature. The RNA was extracted with acid phenol pH 4.7, precipitated using 0.3 M sodium acetate and 2.5 volumes of ethanol, resuspended in RNase-free water, and DNase treated at 37°C. RNA extraction and precipitation were repeated as before. The mRNA was then isolated using the Qiagen Oligotex kit  
5 (QIAGEN, Inc., Chatsworth, CA) and used to construct the cDNA library.

The mRNA was handled according to the recommended protocols in the SuperScript plasmid system (Catalog #18248-013, GIBCO-BRL). cDNA synthesis was initiated with a NotI-oligo d(T) primer. Double stranded cDNA was blunted, ligated to EcoRI adaptors, digested with NotI, fractionated on a Sepharose CL4B column (Catalog  
10 #275105-01; Pharmacia), and those cDNAs exceeding 400 bp were ligated into the NotI and EcoRI sites of the pINCY 1 vector (Incyte). The plasmid pINCY 1 was subsequently transformed into DH5 $\alpha$ <sup>TM</sup> competent cells (Catalog #18258-012; GIBCO-BRL).

## **II Isolation and Sequencing of cDNA Clones**

15 Plasmid DNA was released from the cells and purified using the REAL Prep 96 plasmid kit (Catalog #26173; QIAGEN, Inc.). The recommended protocol was employed except for the following changes: 1) the bacteria were cultured in 1 ml of sterile Terrific Broth (Catalog #22711, GIBCO-BRL) with carbenicillin at 25 mg/l and glycerol at 0.4%; 2) after inoculation, the cultures were incubated for 19 hours and at the end of incubation, the  
20 cells were lysed with 0.3 ml of lysis buffer; and 3) following isopropanol precipitation, the plasmid DNA pellet was resuspended in 0.1 ml of distilled water. After the last step in the protocol, samples were transferred to a 96-well block for storage at 4° C.

The cDNAs were sequenced by the method of Sanger et al. (1975, J. Mol. Biol. 94:441f), using a Hamilton Micro Lab 2200 (Hamilton, Reno, NV) in combination with  
25 Peltier Thermal Cyclers (PTC200 from MJ Research, Watertown, MA) and Applied Biosystems 377 DNA Sequencing Systems.

## **III. Homology Searching of cDNA Clones and Their Deduced Proteins**

The nucleotide sequences and/or amino acid sequences of the Sequence Listing  
30 were used to query sequences in the GenBank, SwissProt, BLOCKS, and Pima II databases. These databases, which contain previously identified and annotated sequences, were searched for regions of homology using BLAST (Basic Local Alignment Search

Tool). (See, e.g., Altschul, S.F. (1993) J. Mol. Evol 36:290-300; and Altschul et al. (1990) J. Mol. Biol. 215:403-410.)

BLAST produced alignments of both nucleotide and amino acid sequences to determine sequence similarity. Because of the local nature of the alignments, BLAST was especially useful in determining exact matches or in identifying homologs which may be of prokaryotic (bacterial) or eukaryotic (animal, fungal, or plant) origin. Other algorithms could have been used when dealing with primary sequence patterns and secondary structure gap penalties. (See, e.g., Smith, T. et al. (1992) Protein Engineering 5:35-51.) The sequences disclosed in this application have lengths of at least 49 nucleotides and have no more than 12% uncalled bases (where N is recorded rather than A, C, G, or T).

The BLAST approach searched for matches between a query sequence and a database sequence. BLAST evaluated the statistical significance of any matches found, and reported only those matches that satisfy the user-selected threshold of significance. In this application, threshold was set at  $10^{-25}$  for nucleotides and  $10^{-8}$  for peptides.

Incyte nucleotide sequences were searched against the GenBank databases for primate (pri), rodent (rod), and other mammalian sequences (mam), and deduced amino acid sequences from the same clones were then searched against GenBank functional protein databases, mammalian (mamp), vertebrate (vrtp), and eukaryote (eukp), for homology.

20

#### IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; and Ausubel, F.M. et al. *supra*, ch. 4 and 16.)

Analogous computer techniques applying BLAST are used to search for identical or related molecules in nucleotide databases such as GenBank or LIFESEQ™ database (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or homologous.

The basis of the search is the product score, which is defined as:

$$\% \text{ sequence identity} \times \% \text{ maximum BLAST score}$$

100

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Homologous molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analysis are reported as a list of libraries in which the transcript encoding MSP occurs. Abundance and percent abundance are also reported. Abundance directly reflects the number of times a particular transcript is represented in a cDNA library, and percent abundance is abundance divided by the total number of sequences examined in the cDNA library.

#### V. Extension of MSP Encoding Polynucleotides

The nucleic acid sequence of Incyte Clones 77138, 1381884, 1457779, 1481261, 1794154, and 1737775 were used to design oligonucleotide primers for extending partial nucleotide sequences to full length. For each nucleic acid sequence, one primer was synthesized to initiate extension of an antisense polynucleotide, and the other was synthesized to initiate extension of a sense polynucleotide. Primers were used to facilitate the extension of the known sequence "outward" generating amplicons containing new unknown nucleotide sequence for the region of interest. The initial primers were designed from the cDNA using OLIGO 4.06 (National Biosciences, Plymouth, MN), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries (GIBCO/BRL) were used to extend the sequence. If more than one extension is necessary or desired, additional sets of primers are designed to further extend the known region.

High fidelity amplification was obtained by following the instructions for the XL-PCR kit (Perkin Elmer) and thoroughly mixing the enzyme and reaction mix. PCR was performed using the Peltier Thermal Cycler (PTC200; M.J. Research, Watertown, MA), beginning with 40 pmol of each primer and the recommended concentrations of all other

components of the kit, with the following parameters:

	Step 1	94° C for 1 min (initial denaturation)
	Step 2	65° C for 1 min
	Step 3	68° C for 6 min
5	Step 4	94° C for 15 sec
	Step 5	65° C for 1 min
	Step 6	68° C for 7 min
	Step 7	Repeat steps 4 through 6 for an additional 15 cycles
	Step 8	94° C for 15 sec
10	Step 9	65° C for 1 min
	Step 10	68° C for 7:15 min
	Step 11	Repeat steps 8 through 10 for an additional 12 cycles
	Step 12	72° C for 8 min
15	Step 13	4° C (and holding)

A 5  $\mu$ l to 10  $\mu$ l aliquot of the reaction mixture was analyzed by electrophoresis on a low concentration (about 0.6% to 0.8%) agarose mini-gel to determine which reactions were successful in extending the sequence. Bands thought to contain the largest products were excised from the gel, purified using QIAQuick™ (QIAGEN Inc., Chatsworth, CA), and trimmed of overhangs using Klenow enzyme to facilitate religation and cloning.

After ethanol precipitation, the products were redissolved in 13  $\mu$ l of ligation buffer, 1  $\mu$ l T4-DNA ligase (15 units) and 1  $\mu$ l T4 polynucleotide kinase were added, and the mixture was incubated at room temperature for 2 to 3 hours, or overnight at 16° C. Competent *E. coli* cells (in 40  $\mu$ l of appropriate media) were transformed with 3  $\mu$ l of ligation mixture and cultured in 80  $\mu$ l of SOC medium. (See, e.g., Sambrook, *supra*, Appendix A, p. 2.) After incubation for one hour at 37° C, the *E. coli* mixture was plated on Luria Bertani (LB) agar (See, e.g., Sambrook, *supra*, Appendix A, p. 1) containing 2x Carb. The following day, several colonies were randomly picked from each plate and cultured in 150  $\mu$ l of liquid LB/2x Carb medium placed in an individual well of an appropriate commercially-available sterile 96-well microtiter plate. The following day, 5  $\mu$ l of each overnight culture was transferred into a non-sterile 96-well plate and, after dilution 1:10 with water, 5  $\mu$ l from each sample was transferred into a PCR array.

For PCR amplification, 18  $\mu$ l of concentrated PCR reaction mix (3.3x) containing 4 units of rTth DNA polymerase, a vector primer, and one or both of the gene specific primers used for the extension reaction were added to each well. Amplification was performed using the following conditions:

- |   |        |  |
|---|--------|--|
|   | Step 1 | 94° C for 60 sec                                     |
|   | Step 2 | 94° C for 20 sec                                     |
|   | Step 3 | 55° C for 30 sec                                     |
|   | Step 4 | 72° C for 90 sec                                     |
| 5 | Step 5 | Repeat steps 2 through 4 for an additional 29 cycles |
|   | Step 6 | 72° C for 180 sec                                    |
|   | Step 7 | 4° C (and holding)                                   |

Aliquots of the PCR reactions were run on agarose gels together with molecular  
 10 weight markers. The sizes of the PCR products were compared to the original partial  
 cDNAs, and appropriate clones were selected, ligated into plasmid, and sequenced.

In like manner, the nucleotide sequences of SEQ ID NO:6, SEQ ID NO:7, SEQ  
 ID NO:8, SEQ ID NO:9, SEQ ID NO:10, and SEQ ID NO:11, are used to obtain 5'  
 regulatory sequences using the procedure above, oligonucleotides designed for 5'  
 15 extension, and an appropriate genomic library.

#### **VI. Labeling and Use of Individual Hybridization Probes**

Hybridization probes derived from SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8,  
 SEQ ID NO:9, SEQ ID NO:10, and SEQ ID NO:11, are employed to screen cDNAs,  
 20 genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of  
 about 20 base pairs, is specifically described, essentially the same procedure is used with  
 larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software  
 such as OLIGO 4.06 (National Biosciences) and labeled by combining 50 pmol of each  
 oligomer, 250  $\mu$ Ci of [ $\gamma$ -<sup>32</sup>P] adenosine triphosphate (Amersham, Chicago, IL), and T4  
 25 polynucleotide kinase (DuPont NEN®, Boston, MA). The labeled oligonucleotides are  
 substantially purified using a Sephadex G-25 superfine resin column (Pharmacia &  
 Upjohn, Kalamazoo, MI). An aliquot containing 10<sup>7</sup> counts per minute of the labeled  
 probe is used in a typical membrane-based hybridization analysis of human genomic DNA  
 digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or  
 30 Pvu II (DuPont NEN, Boston, MA).

The DNA from each digest is fractionated on a 0.7 percent agarose gel and  
 transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham, NH).  
 Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots  
 are sequentially washed at room temperature under increasingly stringent conditions up to

0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMAT ART<sup>TM</sup> film (Kodak, Rochester, NY) is exposed to the blots to film for several hours, hybridization patterns are compared visually.

## 5 VII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using or thermal, UV, mechanical, or chemical bonding procedures, or a  
10 vacuum system. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the  
15 scanned images.

In another alternative, full-length cDNAs or Expressed Sequence Tags (ESTs) comprise the elements of the microarray. Full-length cDNAs or ESTs corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate,  
20 e.g., a glass slide. The cDNA is fixed to the slide using, e.g., U.V. cross-linking followed, by thermal and chemical and subsequent drying. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; and Shalon, D. et al. (1996) Genome Res. 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

Probe sequences for microarrays may be selected by screening a large number of  
25 clones from a variety of cDNA libraries in order to find sequences with conserved protein motifs common to genes coding for signal sequence containing polypeptides. In one embodiment, sequences identified from cDNA libraries, are analyzed to identify those gene sequences with conserved protein motifs using an appropriate analysis program, e.g.,  
30 the Block 2 Bioanalysis Program (Incyte, Palo Alto, CA). This motif analysis program, based on sequence information contained in the Swiss-Prot Database and PROSITE, is a method of determining the function of uncharacterized proteins translated from genomic or

cDNA sequences. (See, e.g., Bairoch, A. et al. (1997) *Nucleic Acids Res.* 25:217-221; and Attwood, T. K. et al. (1997) *J. Chem. Inf. Comput. Sci.* 37:417-424.) PROSITE may be used to identify functional or structural domains that cannot be detected using conserved motifs due to extreme sequence divergence. The method is based on weight matrices.

- 5 Motifs identified by this method are then calibrated against the SWISS-PROT database in order to obtain a measure of the chance distribution of the matches.

In another embodiment, Hidden Markov models (HMMs) may be used to find shared motifs, specifically consensus sequences. (See, e.g., Pearson, W.R. and D.J. Lipman (1988) *Proc. Natl. Acad. Sci.* 85:2444-2448; and Smith, T.F. and M.S. Waterman  
10 (1981) *J. Mol. Biol.* 147:195-197.) HMMs were initially developed to examine speech recognition patterns, but are now being used in a biological context to analyze protein and nucleic acid sequences as well as to model protein structure. (See, e.g., Krogh, A. et al. (1994) *J. Mol. Biol.* 235:1501-1531; and Collin, M. et al. (1993) *Protein Sci.* 2:305-314.) HMMs have a formal probabilistic basis and use position-specific scores for amino acids  
15 or nucleotides. The algorithm continues to incorporate information from newly identified sequences to increase its motif analysis capabilities.

### **VIII. Complementary Polynucleotides**

Sequences complementary to the MSP-encoding sequences, or any parts thereof,  
20 are used to detect, decrease, or inhibit expression of naturally occurring MSP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using Oligo 4.06 software and the coding sequence of MSP. To inhibit transcription, a complementary oligonucleotide is designed from the most  
25 unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the MSP-encoding transcript.

### **IX. Expression of MSP**

30 Expression of MSP is accomplished by subcloning the cDNA into an appropriate vector and transforming the vector into host cells. This vector contains an appropriate promoter, e.g.,  $\beta$ -galactosidase, upstream of the cloning site, operably associated with the

cDNA of interest. (See, e.g., Sambrook, *supra*, pp. 404-433; and Rosenberg, M. et al. (1983) *Methods Enzymol.* 101:123-138.)

Induction of an isolated, transformed bacterial strain with isopropyl beta-D-thiogalactopyranoside (IPTG) using standard methods produces a fusion protein which  
5 consists of the first 8 residues of  $\beta$ -galactosidase, about 5 to 15 residues of linker, and the full length protein. The signal residues direct the secretion of MSP into bacterial growth media which can be used directly in the following assay for activity.

#### **X. Demonstration of MSP Activity**

10 Given the chemical and structural similarity between the MSP and other members of the membrane spanning protein families, MSP is identified as a new member of the membrane spanning proteins and is presumed to be involved in the regulation of cell growth.

To demonstrate that increased levels of MSP expression correlates with decreased  
15 cell motility and increased cell proliferation, expression vectors encoding MSP are electroporated into highly motile cell lines, such as U-937 (ATCC CRL 1593), HEL 92.1.7 (ATCC TIB 180) and MAC10, and the motility of the electroporated and control cells are compared. Methods for the design and construction of an expression vector capable of expressing MSP in the desired mammalian cell line(s) chosen are well known to the art.  
20 Assays for examining the motility of cells in culture are known to the art (cf Miyake, M. et al. (1991) *J. Exp. Med.* 174:1347-1354 and Ikeyama, S. et al. (1993) *J. Exp. Med.* 177:1231-1237). Increasing the level of MSP in highly motile cell lines by transfection with an MSP expression vector inhibits or reduces the motility of these cell lines, and the amount of this inhibition is proportional to the activity of MSP in the assay.

25

#### **XI. Production of MSP Specific Antibodies**

MSP substantially purified using PAGE electrophoresis (see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols. The MSP amino  
30 acid sequence is analyzed using DNASTAR software (DNASTAR Inc) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of



appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel et al. supra, ch. 11.)

Typically, the oligopeptides are 15 residues in length, and are synthesized using an Applied Biosystems Peptide Synthesizer Model 431A using fmoc-chemistry and  
5 coupled to KLH (Sigma, St. Louis, MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel et al. supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide activity, for example, by binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and  
10 reacting with radio-iodinated goat anti-rabbit IgG.

## **XII. Purification of Naturally Occurring MSP Using Specific Antibodies**

Naturally occurring or recombinant MSP is substantially purified by immunoaffinity chromatography using antibodies specific for MSP. An immunoaffinity  
15 column is constructed by covalently coupling anti-MSP antibody to an activated chromatographic resin, such as CNBr-activated Sepharose (Pharmacia & Upjohn). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing MSP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of MSP (e.g.,  
20 high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/MSP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and MSP is collected.

## **XIII. Identification of Molecules Which Interact with MSP**

25 MSP, or biologically active fragments thereof, are labeled with  $^{125}\text{I}$  Bolton-Hunter reagent. (See, e.g., Bolton et al. (1973) Biochem. J. 133:529.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled MSP, washed, and any wells with labeled MSP complex are assayed. Data obtained using different concentrations of MSP are used to calculate values for the number, affinity, and  
30 association of MSP with the candidate molecules.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and

spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in  
5 molecular biology or related fields are intended to be within the scope of the following claims.

What is claimed is:

1. A substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, or fragments thereof.  
5
2. A substantially purified variant having at least 90% amino acid identity to the amino acid sequence of claim 1.
3. An isolated and purified polynucleotide having a nucleic acid sequence  
10 selected from the group consisting of SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, and SEQ ID NO:11.
4. A microarray containing at least a fragment of at least one of the polynucleotides encoding an MSP of claim 1.  
15
5. An isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide of claim 3.
6. An isolated and purified polynucleotide which hybridizes under stringent  
20 conditions to the polynucleotide of claim 3.
7. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide sequence of claim 3.
- 25 8. An isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, and fragments thereof.
9. An isolated and purified polynucleotide variant having at least 90%  
30 polynucleotide sequence identity to the polynucleotide of claim 8.
10. An isolated and purified polynucleotide having a sequence which is

complementary to the polynucleotide of claim 8.

11. An expression vector containing at least a fragment of the polynucleotide of claim 3.
- 5
12. A host cell containing the expression vector of claim 11.
13. A method for producing a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, and fragments thereof, the method comprising the steps of:
- 10
- a) culturing the host cell of claim 12 under conditions suitable for the expression of the polypeptide; and
- b) recovering the polypeptide from the host cell culture.
- 15
14. A pharmaceutical composition comprising the polypeptide of claim 1 in conjunction with a suitable pharmaceutical carrier.
15. A purified antibody which specifically binds to the polypeptide of claim 1.
- 20
16. A purified agonist of the polypeptide of claim 1.
17. A purified antagonist of the polypeptide of claim 1.
18. A method for treating or preventing a neoplastic disorder, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 17.
- 25
19. A method for treating or preventing an immunological disorder, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 17.
- 30
20. A method for treating or preventing a reproductive disorder, the method

comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 17.

21. A method for detecting a polynucleotide encoding the polypeptide  
5 comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, and fragments thereof in a biological sample, the method comprising the steps of:

- (a) hybridizing the polynucleotide of claim 7 to at least one of the  
nucleic acids in the biological sample, thereby forming a hybridization complex;  
10 and  
(b) detecting the hybridization complex, wherein the presence of the  
hybridization complex correlates with the presence of the polynucleotide encoding  
the polypeptide in the biological sample.

15 22. The method of claim 21 wherein the nucleic acids of the biological sample are amplified by the polymerase chain reaction prior to hybridization.

## SEQUENCE LISTING

&lt;110&gt; INCYTE PHARMACEUTICALS, INC.

TANG, Y. Tom

BANDMAN, Olga

LAL, Preeti

HILLMAN, Jennifer L.

YUE, Henry

CORLEY, Neil C.

GUEGLER, Karl J.

KASER, Matthew R.

BAUGHN, Mariah R.

SHAH, Purvi

&lt;120&gt; HUMAN MEMBRANE SPANNING PROTEINS

&lt;130&gt; PF-0489 PCT

&lt;140&gt; To Be Assigned

&lt;141&gt; Herewith

&lt;150&gt; 09/039,064

&lt;151&gt; 1998-03-13

&lt;160&gt; 11

&lt;170&gt; PERL PROGRAM

&lt;210&gt; 1

&lt;211&gt; 238

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt; -

&lt;223&gt; 077138

&lt;400&gt; 1

Met	Leu	Thr	Pro	Leu	Gly	Lys	Phe	Ser	Thr	Ala	Lys	Phe	Ala	Val
1				5					10					15
Arg	Leu	Pro	Arg	Val	Trp	Glu	Ala	Arg	Ala	Pro	Ser	Leu	Ser	Gly
			20						25					30
Ala	Pro	Ala	Pro	Thr	Pro	Pro	Ala	Pro	Pro	Pro	Ser	Arg	Ser	Ser
			35						40					45
Arg	Leu	Gly	Leu	Trp	Pro	Arg	Cys	Phe	Leu	Ile	Phe	Pro	Gln	Leu
			50						55					60
Arg	Ile	Leu	Leu	Leu	Gly	Pro	Gln	Glu	Ser	Asn	Asn	Ser	Thr	Gly
			65						70					75
Thr	Met	Trp	Lys	Val	Ser	Ala	Leu	Leu	Phe	Val	Leu	Gly	Ser	Ala
			80						85					90
Ser	Leu	Trp	Val	Leu	Ala	Glu	Gly	Ala	Ser	Thr	Gly	Gln	Pro	Glu
			95						100					105
Asp	Asp	Thr	Glu	Thr	Thr	Gly	Leu	Glu	Gly	Gly	Val	Ala	Met	Pro
			110						115					120
Gly	Ala	Glu	Asp	Asp	Val	Val	Thr	Pro	Gly	Thr	Ser	Glu	Asp	Arg
			125						130					135

```

Tyr Lys Ser Gly Leu Thr Thr Leu Val Ala Thr Ser Val Asn Ser
      140                      145                      150
Val Thr Gly Ile Arg Ile Glu Asp Leu Pro Thr Ser Glu Ser Thr
      155                      160                      165
Val His Ala Gln Glu Gln Ser Pro Ser Ala Thr Ala Ser Asn Val
      170                      175                      180
Ala Thr Ser His Ser Thr Glu Lys Val Asp Gly Asp Thr Gln Thr
      185                      190                      195
Thr Val Glu Lys Asp Gly Leu Ser Thr Val Thr Leu Val Gly Ile
      200                      205                      210
Ile Val Gly Val Leu Leu Ala Ile Gly Phe Ile Gly Gly Ile Ile
      215                      220                      225
Val Val Val Met Arg Lys Met Ser Gly Arg Tyr Ser Pro
      230                      235

```

<210> 2  
 <211> 83  
 <212> PRT  
 <213> Homo sapiens

<220> -  
 <223> 1381884

```

<400> 2
Met Met Val Thr Val Val Leu His Leu Ala Phe Gly Gln Arg Leu
  1          5          10          15
Ile Asp Phe Ser His His His Thr Val Ser Asn Ile Leu Phe His
      20          25          30
Cys Ala Leu Ile Leu Cys Lys Lys Glu Pro Gly Ala Lys Ile Ser
      35          40          45
Arg Gly Ala Lys Gly Lys Lys Glu Glu Lys Gln Glu Ala Gly Lys
      50          55          60
Glu Gly Thr Ala Pro Ser Glu Asn Gly Glu Thr Lys Ala Glu Glu
      65          70          75
Val Leu Ser Ile Asn Thr Ser His
      80

```

<210> 3  
 <211> 495  
 <212> PRT  
 <213> Homo sapiens

<220> -  
 <223> 1457779

```

<400> 3
Met Gly Leu Ser Arg Lys Glu Gln Val Phe Leu Ala Leu Leu Gly
  1          5          10          15
Ala Ser Gly Val Ser Gly Leu Thr Ala Leu Ile Leu Leu Leu Val
      20          25          30
Glu Ala Thr Ser Val Leu Leu Pro Thr Asp Ile Lys Phe Gly Ile
      35          40          45

```

Val Phe Asp Ala Gly Ser Ser His Thr Ser Leu Phe Leu Tyr Gln	50	55	60
Trp Pro Ala Asn Lys Glu Asn Gly Thr Gly Val Val Ser Gln Ala	65	70	75
Leu Ala Cys Gln Val Glu Gly Pro Gly Ile Ser Ser Tyr Thr Ser	80	85	90
Asn Ala Ala Gln Ala Gly Glu Ser Leu Gln Gly Cys Leu Glu Glu	95	100	105
Ala Leu Val Leu Ile Pro Glu Ala Gln His Arg Lys Thr Pro Thr	110	115	120
Phe Leu Gly Ala Thr Ala Gly Met Arg Leu Leu Ser Arg Lys Asn	125	130	135
Ser Ser Gln Ala Arg Asp Ile Phe Ala Ala Val Thr Gln Val Leu	140	145	150
Gly Arg Ser Pro Val Asp Phe Trp Gly Ala Glu Leu Leu Ala Gly	155	160	165
Gln Ala Glu Gly Ala Phe Gly Trp Ile Thr Val Asn Tyr Gly Leu	170	175	180
Gly Thr Leu Val Lys Tyr Ser Phe Thr Gly Glu Trp Ile Gln Pro	185	190	195
Pro Glu Glu Met Leu Val Gly Ala Leu Asp Met Gly Gly Ala Ser	200	205	210
Thr Gln Ile Thr Phe Val Pro Gly Gly Pro Ile Leu Asp Lys Ser	215	220	225
Thr Gln Ala Asp Phe Arg Leu Tyr Gly Ser Asp Tyr Ser Val Tyr	230	235	240
Thr His Ser Tyr Leu Cys Phe Gly Arg Asp Gln Met Leu Ser Arg	245	250	255
Leu Leu Val Gly Leu Val Gln Ser Arg Pro Ala Ala Leu Leu Arg	260	265	270
His Pro Cys Tyr Leu Ser Gly Tyr Gln Thr Thr Leu Ala Leu Gly	275	280	285
Pro Leu Tyr Glu Ser Pro Cys Val His Ala Thr Pro Pro Leu Ser	290	295	300
Leu Pro Gln Asn Leu Thr Val Glu Gly Thr Gly Asn Pro Gly Ala	305	310	315
Cys Val Ser Ala Ile Arg Glu Leu Phe Asn Phe Ser Ser Cys Gln	320	325	330
Gly Gln Glu Asp Cys Ala Phe Asp Gly Val Tyr Gln Pro Pro Leu	335	340	345
Arg Gly Gln Phe Tyr Ala Phe Ser Asn Phe Tyr Tyr Thr Phe His	350	355	360
Phe Leu Asn Leu Thr Ser Arg Gln Pro Leu Ser Thr Val Asn Ala	365	370	375
Thr Ile Trp Glu Phe Cys Gln Arg Pro Trp Lys Leu Val Glu Ala	380	385	390
Ser Tyr Pro Gly Gln Asp Arg Trp Leu Arg Asp Tyr Cys Ala Ser	395	400	405
Gly Leu Tyr Ile Leu Thr Leu Leu His Glu Gly Tyr Gly Phe Ser	410	415	420
Glu Glu Thr Trp Pro Ser Leu Glu Phe Arg Lys Gln Ala Gly Gly	425	430	435
Val Asp Ile Gly Trp Thr Leu Gly Tyr Met Leu Asn Leu Thr Gly	440	445	450
Met Ile Pro Ala Asp Ala Pro Ala Gln Trp Arg Ala Glu Ser Tyr	455	460	465
Gly Val Trp Val Ala Lys Val Val Phe Met Val Leu Ala Leu Val	470	475	480



Ala Val Val Gly Ala Ala Leu Val Gln Leu Phe Trp Leu Gln Asp  
 485 490 495

<210> 4  
 <211> 156  
 <212> PRT  
 <213> Homo sapiens

<220> -  
 <223> 1481261

<400> 4  
 Met Leu Ser Ala Arg Asp Arg Arg Asp Arg His Pro Glu Glu Gly  
 1 5 10 15  
 Val Val Ala Glu Leu Gln Gly Phe Ala Val Asp Lys Ala Phe Leu  
 20 25 30  
 Thr Ser His Lys Gly Ile Leu Leu Glu Thr Glu Leu Ala Leu Thr  
 35 40 45  
 Leu Ile Ile Phe Ile Cys Phe Thr Ala Ser Ile Ser Ala Tyr Met  
 50 55 60  
 Ala Ala Ala Leu Leu Glu Phe Phe Ile Thr Leu Ala Phe Leu Phe  
 65 70 75  
 Leu Tyr Ala Thr Gln Tyr Tyr Gln Arg Phe Asp Arg Ile Asn Trp  
 80 85 90  
 Pro Cys Leu Asp Phe Leu Arg Cys Val Ser Ala Ile Ile Ile Phe  
 95 100 105  
 Leu Val Val Ser Phe Ala Ala Val Thr Ser Arg Asp Gly Ala Ala  
 110 115 120  
 Ile Ala Ala Phe Val Phe Gly Ile Ile Leu Val Ser Ile Phe Ala  
 125 130 135  
 Tyr Asp Ala Phe Lys Ile Tyr Arg Thr Glu Met Ala Pro Gly Ala  
 140 145 150  
 Ser Gln Gly Asp Gln Gln  
 155

<210> 5  
 <211> 950  
 <212> PRT  
 <213> Homo sapiens

<220> -  
 <223> 1794154

<400> 5  
 Met Thr Trp Arg Met Gly Pro Arg Phe Thr Met Leu Leu Ala Met  
 1 5 10 15  
 Trp Leu Val Cys Gly Ser Glu Pro His Pro His Ala Thr Ile Arg  
 20 25 30  
 Gly Ser His Gly Gly Arg Lys Val Pro Leu Val Ser Pro Asp Ser  
 35 40 45  
 Ser Arg Pro Ala Arg Phe Leu Arg His Thr Gly Arg Ser Arg Gly  
 50 55 60

Ile	Glu	Arg	Ser	Thr	Leu	Glu	Glu	Pro	Asn	Leu	Gln	Pro	Leu	Gln
					65				70					75
Arg	Arg	Arg	Ser	Val	Pro	Val	Leu	Arg	Leu	Ala	Arg	Pro	Thr	Glu
					80				85					90
Pro	Pro	Ala	Arg	Ser	Asp	Ile	Asn	Gly	Ala	Ala	Val	Arg	Pro	Glu
					95				100					105
Gln	Arg	Pro	Ala	Ala	Arg	Gly	Ser	Pro	Arg	Glu	Met	Ile	Arg	Asp
					110				115					120
Glu	Gly	Ser	Ser	Ala	Arg	Ser	Arg	Met	Leu	Arg	Phe	Pro	Ser	Gly
					125				130					135
Ser	Ser	Ser	Pro	Asn	Ile	Leu	Ala	Ser	Phe	Ala	Gly	Lys	Asn	Arg
					140				145					150
Val	Trp	Val	Ile	Ser	Ala	Pro	His	Ala	Ser	Glu	Gly	Tyr	Tyr	Arg
					155				160					165
Leu	Met	Met	Ser	Leu	Leu	Lys	Asp	Asp	Val	Tyr	Cys	Glu	Leu	Ala
					170				175					180
Glu	Arg	His	Ile	Gln	Gln	Ile	Val	Leu	Phe	His	Gln	Ala	Gly	Glu
					185				190					195
Glu	Gly	Gly	Lys	Val	Arg	Arg	Ile	Thr	Ser	Glu	Gly	Gln	Ile	Leu
					200				205					210
Glu	Gln	Pro	Leu	Asp	Pro	Ser	Leu	Ile	Pro	Lys	Leu	Met	Ser	Phe
					215				220					225
Leu	Lys	Leu	Glu	Lys	Gly	Lys	Phe	Gly	Met	Val	Leu	Leu	Lys	Lys
					230				235					240
Thr	Leu	Gln	Val	Glu	Glu	Arg	Tyr	Pro	Tyr	Pro	Val	Arg	Leu	Glu
					245				250					255
Ala	Met	Tyr	Glu	Val	Ile	Asp	Gln	Gly	Pro	Ile	Arg	Arg	Ile	Glu
					260				265					270
Lys	Ile	Arg	Gln	Lys	Gly	Phe	Val	Gln	Lys	Cys	Lys	Ala	Ser	Gly
					275				280					285
Val	Glu	Gly	Gln	Val	Val	Ala	Glu	Gly	Asn	Asp	Gly	Gly	Gly	Gly
					290				295					300
Ala	Gly	Arg	Pro	Ser	Leu	Gly	Ser	Glu	Lys	Lys	Lys	Glu	Asp	Pro
					305				310					315
Arg	Arg	Ala	Gln	Val	Pro	Pro	Thr	Arg	Glu	Ser	Arg	Val	Lys	Val
					320				325					330
Leu	Arg	Lys	Leu	Ala	Ala	Thr	Ala	Pro	Ala	Leu	Pro	Gln	Pro	Pro
					335				340					345
Ser	Thr	Pro	Arg	Ala	Thr	Thr	Leu	Pro	Pro	Ala	Pro	Ala	Thr	Thr
					350				355					360
Val	Thr	Arg	Ser	Thr	Ser	Arg	Ala	Val	Thr	Val	Ala	Ala	Arg	Pro
					365				370					375
Met	Thr	Thr	Thr	Ala	Phe	Pro	Thr	Thr	Gln	Arg	Pro	Trp	Thr	Pro
					380				385					390
Ser	Pro	Ser	His	Arg	Pro	Pro	Thr	Thr	Thr	Glu	Val	Ile	Thr	Ala
					395				400					405
Arg	Arg	Pro	Ser	Val	Ser	Glu	Asn	Leu	Tyr	Pro	Pro	Ser	Arg	Lys
					410				415					420
Asp	Gln	His	Arg	Glu	Arg	Pro	Gln	Thr	Thr	Arg	Arg	Pro	Ser	Lys
					425				430					435
Ala	Thr	Ser	Leu	Glu	Ser	Phe	Thr	Asn	Ala	Pro	Pro	Thr	Thr	Ile
					440				445					450
Ser	Glu	Pro	Ser	Thr	Arg	Ala	Ala	Gly	Pro	Gly	Arg	Phe	Arg	Asp
					455				460					465
Asn	Arg	Met	Asp	Arg	Arg	Glu	His	Gly	His	Arg	Asp	Pro	Asn	Val
					470				475					480
Val	Pro	Gly	Pro	Pro	Lys	Pro	Ala	Lys	Glu	Lys	Pro	Pro	Lys	Lys
					485				490					495

Lys Ala Gln Asp	Lys Ile Leu Ser Asn Glu Tyr Glu Glu Lys Tyr	
500	505	510
Asp Leu Ser Arg	Pro Thr Ala Ser Gln Leu Glu Asp Glu Leu Gln	
515	520	525
Val Gly Asn Val	Pro Leu Lys Lys Ala Lys Glu Ser Lys Lys His	
530	535	540
Glu Lys Leu Glu	Lys Pro Glu Lys Glu Lys Lys Lys Lys Met Lys	
545	550	555
Asn Glu Asn Ala	Asp Lys Leu Leu Lys Ser Glu Lys Gln Met Lys	
560	565	570
Lys Ser Glu Lys	Lys Ser Lys Gln Glu Lys Glu Lys Ser Lys Lys	
575	580	585
Lys Lys Gly Gly	Lys Thr Glu Gln Asp Gly Tyr Gln Lys Pro Thr	
590	595	600
Asn Lys His Phe	Thr Gln Ser Pro Lys Lys Ser Val Ala Asp Leu	
605	610	615
Leu Gly Ser Phe	Glu Gly Lys Arg Arg Leu Leu Leu Ile Thr Ala	
620	625	630
Pro Lys Ala Glu	Asn Asn Met Tyr Val Gln Gln Arg Asp Glu Tyr	
635	640	645
Leu Glu Ser Phe	Cys Lys Met Ala Thr Arg Lys Ile Ser Val Ile	
650	655	660
Thr Ile Phe Gly	Pro Val Asn Asn Ser Thr Met Lys Ile Asp His	
665	670	675
Phe Gln Leu Asp	Asn Glu Lys Pro Met Arg Val Val Asp Asp Glu	
680	685	690
Asp Leu Val Asp	Gln Arg Leu Ile Ser Glu Leu Arg Lys Glu Tyr	
695	700	705
Gly Met Thr Tyr	Asn Asp Phe Phe Met Val Leu Thr Asp Val Asp	
710	715	720
Leu Arg Val Lys	Gln Tyr Tyr Glu Val Pro Ile Thr Met Lys Ser	
725	730	735
Val Phe Asp Leu	Ile Asp Thr Phe Gln Ser Arg Ile Lys Asp Met	
740	745	750
Glu Lys Gln Lys	Lys Glu Gly Ile Val Cys Lys Glu Asp Lys Lys	
755	760	765
Gln Ser Leu Glu	Asn Phe Leu Ser Arg Phe Arg Trp Arg Arg Arg	
770	775	780
Leu Leu Val Ile	Ser Ala Pro Asn Asp Glu Asp Trp Ala Tyr Ser	
785	790	795
Gln Gln Leu Ser	Ala Leu Ser Gly Gln Ala Cys Asn Phe Gly Leu	
800	805	810
Arg His Ile Thr	Ile Leu Lys Leu Leu Gly Val Gly Glu Glu Val	
815	820	825
Gly Gly Val Leu	Glu Leu Phe Pro Ile Asn Gly Ser Ser Val Val	
830	835	840
Glu Arg Glu Asp	Val Pro Ala His Leu Val Lys Asp Ile Arg Asn	
845	850	855
Tyr Phe Gln Val	Ser Pro Glu Tyr Phe Ser Met Leu Leu Val Gly	
860	865	870
Lys Asp Gly Asn	Val Lys Ser Trp Tyr Pro Ser Pro Met Trp Ser	
875	880	885
Met Val Ile Val	Tyr Asp Leu Ile Asp Ser Met Gln Leu Arg Arg	
890	895	900
Gln Glu Met Ala	Ile Gln Gln Ser Leu Gly Met Arg Cys Pro Glu	
905	910	915
Asp Glu Tyr Ala	Gly Tyr Gly Tyr His Ser Tyr His Gln Gly Tyr	
920	925	930

Gln Asp Gly Tyr Gln Asp Asp Tyr Arg His His Glu Ser Tyr His  
 935 940 945  
 His Gly Tyr Pro Tyr  
 950

<210> 6  
 <211> 1823  
 <212> DNA  
 <213> Homo sapiens

<220> -  
 <223> 077138

<400> 6  
 ctagcgcgcc cggacggaga ccaccttgcg gccgaccccg ctcccccgcc tcctcgggag 60  
 agataaatgc tgactccgct cggaaagtcc tcaactgcaa agtttgctgt cgggctgcct 120  
 aggggtctggg aagctcgggc accctccctc tccggggctc ctgctccac ccctccggcc 180  
 cccccaccgt cgcgctcctc caggtcgggc ctgtggcgc ggtgcttttt aattttcccc 240  
 cagctcagaa tcttgctgct cggccccag gagagcaaca actcaacggg aacgatgtgg 300  
 aagggtgtcag ctctgctctt cgttttgga agcgcgtcgc tctgggtcct ggcagaagga 360  
 gccagcacag gccagccaga agatgacact gagactacag gtttggagg cggcgttgcc 420  
 atgccagggtg ccgaagatga tgtggtgact ccaggaacca gcgaagaccg ctataagtct 480  
 ggcttgacaa ctctggtggc aacaagtgtc aacagtgtaa caggcattcg catcgaggat 540  
 ctgccaactt cagaaagcac agtccacgcg caagaacaaa gtccaagcgc cacagcctca 600  
 aacgtggcca ccagtcactc caggagaaaa gtggatggag acacacagac aacagtgtgag 660  
 aaagatgggt tgtcaacagt gacctgggtt ggaatcatag ttgggggtctt actagccatc 720  
 ggcttcattg gtggaatcat cgttggtgtt atgcgaaaaa tgcgaggaa gtactcgccc 780  
 taaagagctg aagggttacg ccctgctgcc aacgtgctta aaaaaagacc gtttctgact 840  
 ctgtgccttg tccctgagct cgtgggagaa gatgacccgt ggaacacttg cctggcccac 900  
 tcagaatcca cgtgacctc tccgcttgcc aaaataaccg aaggaaagac cgttcaccag 960  
 acttggctcc tctaaacatt tgcgtttcaa acatgttttt gaataacat tctataaaag 1020  
 attatttgaa agacaaaatt catagaaaat ggagcaaaac tgtataaact gatttgtaac 1080  
 taacactgga ccattggatc gatattatat gctgtaacca tgtgtctcgc tctgaccatt 1140  
 cttgttattg ttaaaatgca gaggaatctg gaaatattta tatccacgga gtccttggat 1200  
 ccagtgtctac gtcagtaaat agcaccagca ttttgcaatt gctgatctgc tgaatatgtac 1260  
 acattctggg ctagtgttgt ctatctttta aagcctgac tgggtgtgaat aatcaactag 1320  
 gaaatctaaa cttggataac acgtggtgaa caactgcctt tagctgggtcc agattaatca 1380  
 tttcaaagac atccatttta gatcacaagc aggaagtcga tagtctcaa ggcactttgt 1440  
 ttctcccaag taggccacca ggcagcctct agagttgctt taccctaatc cttctccagc 1500  
 catgacttgg tgactctaag cttgctccca cctgccccct ccacttccct cagatgatga 1560  
 ggagccaggg ctaagggggc agccttctct cttcccagtg atgcacatcc ttcacattgg 1620  
 ctgctttgtt ctggaatatg gatattctcag cctggatgcc gaggaagctg ctggatgctt 1680  
 aatggtgcta gaggtcaag tgtgtttgaa accaagagcc agttgtcccc catgcagaaa 1740  
 gaaatcctgt gtgagcctct ggtatgagac ataaaatctg ccagttttat aacactcaaa 1800  
 aaaaaaaaa acacaaaaa aaa 1823

<210> 7  
 <211> 1416  
 <212> DNA  
 <213> Homo sapiens

<220> -  
 <223> 1381884

&lt;400&gt; 7

```

attgtgtttt ccacagaaga gaaagtacac tggactttct gtgcaacctg ttactacatt 60
ttcacagaga ctcatatttg tgcagtgtaa ctcagttgaa acccagcaaa attaggctcc 120
cgtgtctcca taaaggccac catgatggta acggttgtag ttcaccttgc gtttggacag 180
aggctgattg attttagcca tcatcacacc gtgtctaaca ttctctttca ctgtgctttg 240
atcctctgta agaaagaacc tggagcaaag attagcagag gtgctaaagg gaagaaggag 300
gaaaagcagg aagctggaaa ggaaggtact gcaccatctg aaaatgggta aactaaagct 360
gaagaggtag tttccataaa tacctcccac tgattgaatc agtgtcttta aagaaatttc 420
tcaatccttc agccggtgat agcacgttct taatgtctct ttttattgcc tgtaattgta 480
ttgcagatcc acatctctcg ctcaactgtt aatgtctcaa cctccagagg caccaccacc 540
agcacactgt cagtaaaggg gcagattgaa acagtggagag ttaagggtag agtagaaaat 600
tctgcatgtt tgcagtgact agaatacagat agtagtgagg tgggtttttt tttaatcatt 660
atgaagagtg ggagcttgca ggtaaggctt ctgtgggtgt ttgaaaagca gaaagcaata 720
aatgaaacaa agtgtttttg taatatattc ctgccttgct ttcttcaact agagtggaaa 780
taggttttgc agtaaaagct gaaaaaaaaa agaaaacaaa tgttcaaaac tgtgtgtgtt 840
gggtgggtgga atttcctttg cttatagtag ttccagtagt aactatatgt ttttttttcc 900
tttctttttc acaggcacag aaaactgaat ctgtagataa cgagggagaa tgaattgtca 960
tgaaaaattg gggttgattt tatgtatctc ttgggacaac ttttaaaagc tatttttacc 1020
aagtattttg taaatgctaa ttttttagga ctctactagt tggcatacga aaatatataa 1080
ggatggacat tttatcgtct catagtcagt ctttttggaa atttacctca tcctcaagta 1140
aaataaatat cagttaaata ttggaagctg tgtgtaagat tgattcagca ttccatgcac 1200
ttgctttaaa atttagtctt gtgcatactg tgggtgtttt actgtgcata tttgaatttt 1260
tcatgcagtt tttctagagc aataatcagt ggtgcttttg tacctagggt ttagtggatt 1320
ttaatgaaac atggatagtt gtggccacct gctgactatt tgtgggttaa aataaaaagg 1380
ttacttgtct gcagaaaaaa aaaaaaaaaa ggctgt 1416

```

&lt;210&gt; 8

&lt;211&gt; 2261

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt; -

&lt;223&gt; 1457779

&lt;400&gt; 8

```

cctggagcca ggtgcacagc gcatcgcccg aggtgtgtcac cgccctgccc cgcccccccc 60
agctgtcctg gaccagggg caggagagg ctggacgcca ggtgcgcgga cacagaagcg 120
tctaagcaca gcttctctct tgccgctccg ggaagtgggc agccagccca ggagccagta 180
ccacctgcac catggggctg tcccggaaag agcaggtctt cttggccctg ctgggggcct 240
cggggggtctc aggcctcacg gcaactatc tctctctggg ggaggccacc agcgtgtctc 300
tgccacagca catcaagttt gggatcgtgt ttgatgcggg ctctcccac acgtccctct 360
tcctgtatca gtggccggcg aacaaggaga atggcacggg tgtggtcagc caggccctgg 420
cctgccaggt ggaagggcct ggaatctcct cctacacttc taatgctgca caggctgggtg 480
agagcctgca gggctgcttg gaggaggcgc tgggtgctgat ccagaggcc cagcatcgga 540
aaacaccac gttcctgggg gccacggctg gcatgaggtt gctcagccgg aagaacagct 600
ctcaggccag ggacatctt gcagcagtca ccaggtcct gggccggtct cccgtggact 660
tttgggtgac cgagctcctg gccgggcagg ccgaagggtg ctttgggttg atcactgtca 720
actacggctt ggggacgctg gtcaagtact ccttcaactg agaatggatc cagcctccgg 780
aggagatgct ggtgggtgcc ctggacatgg gaggggcctc caccagatc acgttcgtgc 840
ctgggggccc catcttgga aagagcccc aggcggattt tcgcctctac ggctccgact 900
acagcgtcta cactcacagc tacctgtgct ttggacggga ccagatgctg agcaggctcc 960
tcgtggggct ggtgcagagc cgcccggtg ccctgtctcg tcaccctgac tacctcagcg 1020
gctaccagac cacactggcc ctgggcccgc tgtatgagtc acctgtgtc cagccacgc 1080
ccccgctgag cctccccag aacctcacag ttgaaggagc aggcaccct ggagcctgag 1140
tctcagccat ccgggaactt ttcaacttct ccagctgcca gggccaggag gactgcgcct 1200
ttgacggggc ctaccagccc ccgctgcggg gccagttcta tgcttctctc aacttctact 1260
acaccttcca cttcctgaac ctcacctcca ggcagccct gagcacggtc aacgccacca 1320

```

```

tctgggagtt ttgccagagg ccctggaaac tggtaggagg cagctaccct gggcaggacc 1380
gctgggtgcg ggactactgt gcctcaggcc tgtacatcct caccctcctg cacgagggct 1440
acgggttcag cgaggagacc tggcccagcc tcgagttccg aaagcaggcg ggcggtgtgg 1500
acattggctg gacactgggc tacatgctga acctgaccgg gatgatcccg gccgatgcgc 1560
cggctcagtg gcgggcagag agctacggcg tctgggtggc caaagtgggtg ttcattggtgc 1620
tgcccttgtt ggcggtgtgt ggggctgcct tggtaggagg cttctggttg caggactagt 1680
gggaaggcgg aggtgggccc ccacagagcc cacaggcagc tgcgtcccgg atgctggagg 1740
cttcctgagc cctgagcgcc gtggggcctt gctctgtggg tctgcccacg gtcagggtgac 1800
agccacctcc agggcaccgt cagggtgtgt ctggccacag aggctgcatg acctccccctc 1860
ccggcgctcc tcccccaacc tccttccgca actgggcttc cagggccgta ggtgcctttc 1920
tgcacacagg ccgccaggac tcgtggtgtc tccaggctgt gtgactgcag ggccacatgc 1980
tgcctgcaaa cagggaaga ccacggaggc acagggttcc tgctcctgat ggggcctcag 2040
gaggggcgga gaggggtgga agggaggagg ctgccccacc tggacccccg ctctccctgc 2100
tgttgtctga gcagatggat ggagtccagg cctgggggct tctgctgggc cagcccgcc 2160
tcccacacc acttgagggg tgagactgca gtgggggttg tttttattaa aagcatcatg 2220
gacacagcaa aaaaaaaaaa aaaaaaaaaa aaaaaaaag g 2261

```

&lt;210&gt; 9

&lt;211&gt; 1034

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt; -

&lt;223&gt; 1481261

&lt;400&gt; 9

```

aaggggaggg attagagcct ccttctcttc tgcacctccc catgggtctc tagggggctg 60
gtgcaggcag cagcagaggc actctgggca gctgggtgag ggcccatctg ggcaaggccc 120
ccagcgccctg ccttctctcc cggggccctg tgggcaagcc tctgcttca ctttcagggt 180
tctcgaagtg ccttcttgct cctgtctgtt tccccatcct gccagatttc tgtttctctt 240
gtgggctttt tggcagtagg gggctgtgtt ggtgggccc acgaagatgc tcagtgtctg 300
agatcgccgg gaccggcacc ctgaggagg ggtagttgca gagctccagg gcttcgcggt 360
ggacaaggcc ttcctcacct cccacaagg catcctgctg gaaaccgagc tggccctgac 420
cctcatcatc ttcattctgt tcacggcctc catctctgcc tacatggccg cggcgctact 480
ggagtctctc atcacacttg ccttctctct cctctatgcc acccagtact accagcgctt 540
cgaccgaatt aactggccct gtctggactt cctgcgctgt gtcagtgcc tcatcatctt 600
cctggtggtc tcctttgcag ctgtgacctc ccgggaagg gctgccattg ctgcttttgt 660
ttttggcatc atcctggttt ccattcttgc ctatgatgcc ttcaagatct accggactga 720
gatggcaccc ggggcccagcc agggggacca gcagtgactc tggggctacc tggctcctag 780
gccagccag ccagagagga cagtggagcc cagacacgtc tccttgggat tctactagccc 840
ccagcccgcg aaacgccacc ccagccctac acagcagtct ggctgagac gtcactgggg 900
acttatctgt ggagcctggt gctccaggat gtggcttctc atgaagctct ggccagagga 960
ggggaactta ttgggggagg ggggggtggag ggggggaatc tggacctcta agtcattccc 1020
ataattaaaa tacc 1034

```

&lt;210&gt; 10

&lt;211&gt; 4117

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt; -

&lt;223&gt; 1794154

&lt;400&gt; 10

```

taagttaccc cgatgacttg gtttgaagg ggttaaggca ccagtcattc ttttctaaag 60
tgatttatga tgatgtgtgg agtttaaaaa ctttacccca ccccaaagaa cagccctctc 120

```

```

actcctcact gagtccactc tgaacgtgct aaaatgggaa ggaggcgggtg ttttgatgat 180
ctgttaaatt cttagtgaag tttccttgat ttccagtggc tgctgttggt tgagtttggg 240
ttggagcaaa actgaggtag tcctaacatt tctgggactg aatccaggca agagaaaaga 300
gaaaaagaag aagaaaaaga ggaggaaaaa ggtagggaga aataaaggga ggagagaagc 360
acagtgaaaa aaaaaaaaag tcccttttcg acatcacatt cctgtgtttt cctcagcct 420
ggaaaacata tgaatcccag tgcttttacg cccggaaaca aagagactaa gccagactat 480
gggggaaagg gagataagaa ggatcctgga actttaaaaga gggaaagagt gagattcaga 540
aatcgccagg actggacttt aagggacgtc ctgtgtcagc acaagggact ggcacacaca 600
gacacacgag accgaggaga aactgcagac aaatggagat acaaagactt agaaggacag 660
ctcctttcac ctcatcctac ttgtccagaa ggtaaaaaaga cacagccaga aagaaaaggc 720
atcggtcag ctctcagatc aggacaggct gtggatctgt ggcggtactc tgaaagctgg 780
agctgcagca cccccctttt gtattgctca ccctcggtta agagagagag ggctgggagg 840
aaaagtagtt catctaggaa actgtcctgg gaaccaaact tctgatttct tttgcaacct 900
tctgcattcc atctctatga gccaccattg gattacacaa tgacatggag aatgggacct 960
cgtttcacta tgctgttggc catgtggcta gtgtgtggat cagaacccca ccccatgcc 1020
actattagag gcagccacgg aggacggaaa gtgcctttgg tttctccgga cagcagtagg 1080
ccagctcggg ttctgaggca cactgggagg tctcgcgga ttgagagatc cactctggag 1140
gaaccaaaacc ttcagcctct ccagagaagg aggtgtgtgc ccgtgttgag actagctcgc 1200
ccaacagagc cgccagcccg ctcgacatc aatggggccg ccgtgagacc tgagcaaaga 1260
ccagcagcca ggggctctcc gcgtgagatg atcagagatg aggggtcctc agctcgggtca 1320
agaatgttgc gtttcccttc ggggtccagc tctcccaaca tcttgccag ctttgagg 1380
aagaacagag tatgggtcat ctcagccct catgcctcgg aaggctacta ccgctcatg 1440
atgagcctgc tgaaggacga tgtgtactgt gagctggcgg agaggcacat ccaacagatt 1500
gtgctcttcc accaggcagg tgaggaagga ggcaaggtga gaaggatcac cagcgagggc 1560
cagatcctgg agcagccct ggaccctagc ctcatcccta agctgatgag cttcctgaag 1620
ctggagaagg gcaagtttgg catggtgtcg ctgaagaaga cgctgcaggt ggaggagcgc 1680
tatccatata ccgttaggct ggaagccatg tacgaggtca tgcaccaagg ccccatccgt 1740
aggatcgaga agatcaggca gaagggtctt gtccagaaat gtaaggcctc tgggtgtagag 1800
ggccagggtg tggcgagggg gaatgacggg ggagggggag caggaaggcc aagcctgggc 1860
agcgagaaga agaaagagga cccaaggaga gcacaagtcc caccaaccag agagagtcgg 1920
gtgaaggtcc tgagaaaact ggccgccact gcaccagctt tgccccaacc tccctcaacc 1980
cccagagcca ccacccttcc tctgcccc ggcacaacag tgactcggtc cagtcctcgg 2040
gcggtaacag ttgctgcaag acctatgacc accactgcct tccccaccac gcagaggccc 2100
tggaccctcc caccctccca caggccccct acaaccactg aggtgatcac tgccaggaga 2160
ccctcagttt cagagaatct ttaccctcca tcccggaagg atcagcacag ggagaggcca 2220
cagacaacca ggaggcccag caaggccacc agcttgagga gcttcacaaa tgccctccc 2280
accaccatct cagaaccag cacaagggt gctggcccag gccgtttccg ggacaaccgc 2340
atggacaggc ggggaacatg ccaccgagac ccaatgtgg tgccagggtcc tccaagcca 2400
gcaaaggaga aacctccaa aaagaaggcc caggacaaa ttcttagtaa tgagtatgag 2460
gagaagtatg acctcagccg gcctactgcc tctcagctgg aggacgagct gcagggtggg 2520
aatgttcccc ttaaaaaagc aaaggagtct aaaaagcatg aaaagctga gaaaccagag 2580
aaggagaaga aaaaaaagat gaagaatgag aacgcagaca agttacttaa gagtgaagag 2640
caaatgaaga agtctgagaa aaagagcaag caagagaaag agaagagcaa gaagaaaaa 2700
ggaggtaaaa cagaacagga tggctatcag aaaccacca acaaactt cagcagagt 2760
cccaagaagt cagtggccga cctgctggg tctttgaag gcaaacgaag actccttctg 2820
atcactgctc ccaaggctga gaacaatat tatgtgcaac aacgtgatga atatctggaa 2880
agtttctgca agatggctac caggaaaatc tctgtgatca ccatcttcgg cctgtcaac 2940
aacagcacca tgaaaatcga ccactttcag ctagataatg agaagcccat gcgagtgggtg 3000
gatgatgaag acttggtaga ccagcgtctc atcagcgagc tgaggaaaga gtacggaatg 3060
acctacaatg acttcttcat ggtgctaaca gatgtggatc tgagagtcaa gcaatactat 3120
gaggtacca taacaatgaa gtctgtgtt gatctgatcg atactttcca gtcccgaatc 3180
aaagatatgg agaagcagaa gaaggagggc attgtttgca aaggagcaa aaagcagtc 3240
ctggagaact tcctatccag gttccgggtg aggaggagg tgctgggtgat ctctgtcct 3300
aacgatgaag actgggccta ttcacagcag ctctctgccc tcagtgggtca ggcgtgcaat 3360
tttggctctg gccacataac cattctgaag cttttaggcg ttggagagga agttggggga 3420
gtgttagaac tgttccaat taatgggagc tctgtgttg agcgagaaga cgtaccagcc 3480
catttggtga aagacattcg taactatctt caagtgagcc cggagtactt ctccatgctt 3540
ctagtcggaa aagacggaaa tgtcaaatcc tgggtatcct cccaatgtg gtccatggtg 3600

```

```

attgtgtacg atttaattga ttcgatgcaa cttcggagac aggaaatggc gattcagcag 3660
tcaactgggga tgcgctgccc agaagatgag tatgcaggct atggttacca tagttaccac 3720
caaggatacc aggatgggta ccaggatgac tacccgtcac atgagagtta tcaccatgga 3780
tacccttact gagcagaaat atgtaacctt agactcagcc agtttcctct gcagctgcta 3840
aaactacatg tggccagctc cattcttcca cactgcgtac tacatttcct gcctttttct 3900
ttcagtgttt ttctaagact aaataaatag caaactttca cctattcatg agttattatt 3960
gaaacctcaa atcataaaga catttaaaag aattgttttt ctaactggag gggctctagt 4020
gctaaataat agtactgaaa attgatatta ttttcctttt cttatatgaa ggaccttatt 4080
tggcatataa aattttataa aatatgtaaa aaaaaaa 4117

```

<210> 11

<211> 3111

<212> DNA

<213> Homo sapiens

<220> -

<223> 1737775

<400> 11

```

cggctcgagg aaatcacagg gagatgtaca gcaatggggc catttaagag ttctgtgttc 60
atcttgattc ttcaccttct agaagggggc ctgagtaatt cactcattca gctgaacaac 120
aatggctatg aaggcattgt cgttgcaatc gaccccaatg tgccagaaga tgaaacactc 180
attcaacaaa taaaggacat ggtgaccag gcactctctg atctgtttga agctacagga 240
aagcgatttt atttcaaaaa tgttgccatt ttgattcctg aaacatggaa gacaaaggct 300
gactatgtga gacaaaaact tgagacctac aaaaatgctg atgttctggg tgctgagtct 360
actcctccag gtaatgatga accctacact gagcagatgg gcaactgtgg agagaagggt 420
gaaaggatcc acctcactcc tgatttcatt gcaggaaaaa agttagctga atatggacca 480
caaggtaggg catttgtcca tgagtgggct catctacgat ggggagtatt tgacgagtac 540
aataatgatg agaaattcta cttatccaat ggaagaatac aagcagtaag atgttcagca 600
ggtattactg gtacaaatgt agtaaaagaag tgtcagggag gcagctgtta caccaaaaga 660
tgacattcca ataaagtaac aggactctat gaaaaaggat gtgagtttgt tctccaatcc 720
cgccagacgg agaaggcttc tataatgttt gcacaacatg ttgattctat agttgaattc 780
tgtacagAAC aaaaccacaa caaagaagct ccaacaagc aaaatcaaaa atgcaatctc 840
cgaagcacat gggaagtgat ccgtgattct gaggacttta agaaaaccac tcctatgaca 900
acacagccac caaatccac cttctcattg ctgcagattg gacaaagaat tgtgtgttta 960
gtccttgaca aatctggaag catggcgact ggtaaccgcc tcaatcgact gaatcaagca 1020
ggccagcttt tctgtgtgca gacagttagg ctggggctct gggttgggat ggtgacattt 1080
gacagtgtcg cccatgtaca aagtgaactc atacagataa acagtggcag tgacagggac 1140
acactcgcca aaagattacc tgcagcagct tcaggaggga cgtccatctg cagcgggctt 1200
cgatcggcat ttactgtgat taggaagaaa tatccaactg atggatctga aattgtgtctg 1260
ctgacggatg gggaagacaa cactataagt ggggtgcttta acgagggtcaa acaaagtggg 1320
gccatcatcc acacagtgcg tttggggccc tctgcagctc aagaactaga ggagctgtcc 1380
aaaatgacag gaggtttaca gacatatgct tcagatcaag ttcagaacaa tggcctcatt 1440
gatgcttttg gggccctttc atcaggaaat ggagctgtct ctcagcgtct catccagctt 1500
gagagtaagg gattaaccct ccagaacagc cagtggatga atggcacagt gatcgtggac 1560
agcaccgtgg gaaaggacac tttgtttctt atcacctgga caacgcagcc tccccaaatc 1620
cttctctggg atcccagtgg acagaagcaa ggtggctttg tagtggacaa aaacacaaaa 1680
atggcctacc tccaaatccc aggcattgct aagggttgca cttggaaata cagtctgcaa 1740
gcaagctcac aaaccttgac cctgactgtc acgtcccgtg cgtccaatgc taccctgcct 1800
ccaattacag tgacttccaa aacgaacaag gacaccagca aattccccag ccctctggta 1860
gtttatgcaa atattcgcca aggagcctcc ccaatttctc gggccagtgt cacagccctg 1920
attgaatcag tgaatggaaa aacagttacc ttggaactac tggataatgg agcaggtgct 1980
gatgtacta aggatgacgg tgtctactca aggtatttca caacttatga cacgaatggg 2040
agatacagtg taaaagtgcg ggctctggga ggagttaacg cagccagacg gagagtgata 2100
ccccagcaga gtggagcact gtacatacct ggctggattg agaatgatga aatacaatgg 2160
aatccaccaa gacctgaaat taataaggat gatgttcaac acaagcaagt gtgtttcagc 2220
agaacatcct cgggaggctc atttgtggct tctgatgtcc caaatgctcc catacctgat 2280

```



```
ctcttccac ctggccaaat caccgacctg aaggcggaat ttcacggggg cagtctcatt 2340
aatctgactt ggacagctcc tggggatgat tatgaccatg gaacagctca caagtataatc 2400
attcgaataa gtacaagtat tcttgatctc agagacaagt tcaatgaatc tcttcaagtg 2460
aatactactg ctctcatccc aaaggaagcc aactctgagg aagtcttttt gtttaaacca 2520
gaaaacatta cttttgaaaa tggcacagat cttttcattg ctattcaggc tgttgataag 2580
gtcgatctga aatcagaaat atccaacatt gcacgagtat ctttgtttat tcctccacag 2640
actccgccag agacacctag tcctgatgaa acgtctgctc cttgtcctaa tattcatatc 2700
aacagcacca ttcctggcat tcacatttta aaaattatgt ggaagtggat aggagaactg 2760
cagctgtcaa tagcctaggg ctgaattttt gtcagataaa taaaataaat cattcatcct 2820
tttttttgat tataaaatth tctaaaatgt attttagact tcctgtaggg ggcgatatac 2880
taaagtata tagtacattt atactaaatg ttttcctgta gggggcgata tactaaatgt 2940
atthtagact tcctgtaggg ggcgataaaa taaaatgcta aacaactggg tatacatgca 3000
taaaaactat ccattcaaac ccaaaaatth aataatcatt gagtctthta ttaatgaatt 3060
tgaatactag aaagaaacag ggcttgcata aataaatgga agtatgagt t 3111
```



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification<sup>6</sup> : C12N 15/12, C07K 14/705, C12Q 1/68, C07K 16/28, G01N 33/68, A61K 38/16, 35/00</p>	<p>A3</p>	<p>(11) International Publication Number: <b>WO 99/46380</b></p> <p>(43) International Publication Date: 16 September 1999 (16.09.99)</p>
<p>(21) International Application Number: PCT/US99/05073</p> <p>(22) International Filing Date: 9 March 1999 (09.03.99)</p> <p>(30) Priority Data: 09/039,064 13 March 1998 (13.03.98) US</p> <p>(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application US 09/039,064 (CIP) Filed on 13 March 1998 (13.03.98)</p> <p>(71) Applicant (for all designated States except US): INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).</p> <p>(72) Inventors; and (75) Inventors/Applicants (for US only): TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 9054 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #12, Mountain View, CA 94040 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). CORLEY, Neil, C. [US/US]; 1240</p>	<p>Dale Avenue #30, Mountain View, CA 94040 (US). GUEGLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). KASER, Matthew, R. [GB/US]; 4793 Ewing Road, Castro Valley, CA 94546-1017 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). SHAH, Purvi [IN/US]; 1608 Queen Charlotte Drive #5, Sunnyvale, CA 94087 (US).</p> <p>(74) Agents: BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).</p> <p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p> <p><b>Published</b></p> <p><i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p> <p>(88) Date of publication of the international search report: 16 December 1999 (16.12.99)</p>	
<p>(54) Title: HUMAN MEMBRANE SPANNING PROTEINS</p>		
<p>(57) Abstract</p> <p>The invention provides a human membrane spanning proteins (MSPs) and polynucleotides which identify and encode MSPs. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for treating or preventing disorders associated with expression of MSPs.</p>		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/05073

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/12 C07K14/705 C12Q1/68 C07K16/28 G01N33/68  
 A61K38/16 A61K35/00

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C07K C12Q G01N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>ZIMMER G. ET AL.: "Molecular characterization of gp40, a mucin-type glycoprotein from the apical plasma membrane of Madin-Darby canine kidney cells (type I)"            BIOCHEMICAL JOURNAL,            vol. 326, no. 1, 15 August 1997,            pages 99-108, XP002110714            see page 104; figure 4            see page 107, left-hand column, paragraph 2 - right-hand column, paragraph 1; figure 4</p> <p style="text-align: center;">---</p> <p style="text-align: center;">-/--</p>	1-15, 17, 18, 21, 22



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

## \* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"Z" document member of the same patent family

Date of the actual completion of the international search

29 July 1999

Date of mailing of the international search report

03. 11. 1999

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
 NL - 2280 HV Rijswijk  
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
 Fax: (+31-70) 340-3016

Authorized officer

Macchia, G

## INTERNATIONAL SEARCH REPORT

Internat Application No

PCT/US 99/05073

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X 1-15,21, 22	Database EMBL ID HSA225022 er AJ225022  RNA for membrane glycoprotein gp36  entity with Seq.ID:6 nt.163-898 XP002110716 see the whole document paragraph 1; figure 4  ---	
X	MA T. ET AL.: "Evidence against a role of mouse, rat, and two cloned human T1alpha isoforms as a water channel or a regulator of Aquaporin-type water channels" AMERICAN JOURNAL OF RESPIRATORY CELL AND MOLECULAR BIOLOGY, vol. 19, no. 1, January 1998, pages 143-149, XP002110715 see page 145 - paragraph 1; figure 1 see page 146 - paragraph 1; figure 2  ---	1-15,21, 22
P,X	WO 98 45435 A (GENETICS INST INC (US); JACOBS; MCCOY; LAVALLIE; RACIE; MERBERG ET AL) 15 October 1998 see abstract paragraph 1; figure 2 see page 56, line 28-34 - paragraph 1; figure 2 see page 63, line 1 - page 80, line 4; figure 2 Seq.ID:138 is 99% identical to Seq.ID:6 from nt.239 to nt.566 see page 141 - line 4; figure 2  ---	1-22
A	Database EMBL ID HS37E3R Accession number Z60873 19 October 1995 97% identity with Seq.ID:6 nt.1-230 XP002110717 see the whole document line 4; figure 2  ---	
A	WO 95 35505 A (LELAND STANFORD JUNIOR UNIVERSITY (US); SHALON T.D.; BROWN P.O.) 28 December 1995 cited in the application see abstract line 4; figure 2  ---	4
	---	
	-/--	

# INTERNATIONAL SEARCH REPORT

Internal    Application No  
PCT/US 99/05073

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>TASHIRO K. ET AL.: "Signal sequence trap: a cloning strategy for secreted proteins and type I membrane proteins" SCIENCE, vol. 261, 30 July 1993, pages 600-603, XP002911163 see abstract line 4; figure 2 see page 602, middle column - line 4; figure 2</p> <p style="text-align: center;">---</p>	
A	<p>YOKOYAMA-KOBAYASHI M. ET AL.: "A signal sequence detection system using secreted protease activity as an indicator" GENE, vol. 163, no. 2, 3 October 1995, pages 193-196, XP004041983 see abstract line 4; figure 2</p> <p style="text-align: center;">---</p>	
A	<p>LIBERT F. ET AL.: "Selective amplification and cloning of four new members of the G protein-coupled receptor family" SCIENCE, vol. 244, 5 May 1989, pages 569-572, XP002041588</p> <p style="text-align: center;">---</p>	
A	<p>KYTE J. AND DOOLITTLE R.F.: "A simple method for displaying the hydropathic character of a protein" JOURNAL OF MOLECULAR BIOLOGY, vol. 157, no. 1, 5 May 1982, pages 105-132, XP000609692</p> <p style="text-align: center;">---</p>	
A	<p>HARTMANN E. ET AL.: "Predicting the orientation of eukaryotic membrane-spanning proteins" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 86, no. 15, 1 August 1989, pages 5786-5790, XP002002857</p> <p style="text-align: center;">-----</p>	

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 99/05073

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:  
  
see FURTHER INFORMATION sheet PCT/ISA/210
2. ☐ Claims Nos.:  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

see FURTHER INFORMATION sheet, subject 1.

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

# INTERNATIONAL SEARCH REPORT

International Application No. PCT/US 99/05073

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

1. Claims: 1-22 all partially

A polypeptide comprising an aminoacid sequence as in Seq.ID:1 or a variant thereof. A polynucleotide having a nucleic acid sequence as in Seq.ID:6, variants, microarrays, hybridizing, complementary polynucleotides, fragments thereof. An expression vector containing at least a fragment of said polynucleotide, a host cell containing said expression vector, a method for producing said polypeptide by means of said host cell. A pharmaceutical composition comprising said polypeptide. An antibody which specifically binds to said polypeptide. Agonists and antagonists of said polypeptide. Therapeutical applications of said antagonist. Application of said polynucleotide in diagnostics.

2. Claims: 1-22 all partially

As invention 1 but concerning Seq.ID:2 and 7.

3. Claims: 1-22 all partially

As invention 1 but concerning Seq.ID:3 and 8.

4. Claims: 1-22 all partially

As invention 1 but concerning Seq.ID:4 and 9.

5. Claims: 1-22 all partially

As invention 1 but concerning Seq.ID:5 and 10.

6. Claims: 3, 5-12 all partially

A polynucleotide having a nucleic acid sequence as in Seq.ID:11, variants, hybridizing, complementary polynucleotides, fragments thereof. An expression vector containing at least a fragment of said polynucleotide, a host cell containing said expression vector.



## INTERNATIONAL SEARCH REPORT

International Application No. PCT/ US 99/05073

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box 3.

Remark: Although claims 18-20 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

-----

Further defect(s) under Article 17(2)(a):

Continuation of Box 3.

Present claims 16, 17 and dependent claims thereof, relate to an extremely large number of possible compounds. Support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT is to be found, however, for only a very small proportion of the compounds claimed. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Consequently, the search has been carried out for those parts of the claims which appear to be supported and disclosed, namely those parts relating to peptides derived from polypeptide, whose aminoacid sequence is disclosed in Seq.ID:1, specific antibodies and antisenses molecules of the polynucleotide as disclosed in Seq.ID:6.

The applicant's attention is drawn to the fact that claims relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

# INTERNATIONAL SEARCH REPORT

Information on patent family members

Internat. J. Application No

PCT/US 99/05073

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9845435 A	15-10-1998	AU 6956698 A	30-10-1998
WO 9535505 A	28-12-1995	US 5807522 A	15-09-1998
		AT 180570 T	15-06-1999
		AU 709276 B	26-08-1999
		AU 2862995 A	15-01-1996
		CA 2192095 A	28-12-1995
		DE 69509925 D	01-07-1999
		EP 0804731 A	05-11-1997
		EP 0913485 A	06-05-1999
		JP 10503841 T	07-04-1998